THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 60, PART 1 1 January 1948 No. 337

The Propagation of Shock Waves in Steel and Lead

BY D. C. PACK, W. M. EVANS AND H. J. JAMES
Armament Research Department, Ministry of Supply

Communicated by N. F. Mott ; MS. received 18 *April* 1947

ABSTRACT. An investigation is made of the stress system set up by an explosive detonating in contact with a metal surface. An extrapolation from data on the compressibility of steel and lead leads to the conclusion that the shock wave set up by the detonation has an initial velocity in steel which is less, and an initial velocity in lead which is greater, than the velocity of plane elastic waves. The time taken by the fastest pulse to penetrate various lengths of steel and of lead has been measured experimentally, and the results confirm that the plane elastic waves move more quickly for steel ; while for lead the shock wave before damping has a velocity well in excess of that of the elastic waves.

§1. INTRODUCTION

THE propagation of shock waves in metals is a subject which appears to have been but little studied from a fundamental point of view. The behaviour of a shock wave in a metallic medium will depend upon the compressibility of the medium, which in turn depends upon the type of deformation which the medium experiences. The problem is, then, in general, a very difficult one, but it is possible to obtain useful information by restriction to the simpler types of deformation, which permit a simple interpretation. It is clear that the problem is one with a number of important applications, not least in connection with detonation, by which very severe shock waves are set in motion; the movement of the metal tube confining a charge is an obvious example, since it can control the rate at which detonation proceeds in the explosive column.

A series of experiments described below has been carried out with the object of measuring the velocity of the fastest pulse induced in steel and in lead respectively, by detonating a charge in contact with the metal. By plotting the time required for passage through various thicknesses, the initial velocity of the pulse may be found. A study of the stress system set up by the detonation, together with an extrapolation from data published by Bridgman (1931, 1940) on the compressibility of steel and lead, enables a theoretical evaluation to be made of the initial shock wave velocity in each medium. It is found in the case of steel that this velocity is rather less than that of the fastest elastic waves; the latter are, in fact, recorded in the experiments. On the other hand, for lead, the shock wave velocity is greater than the velocity of the elastic waves, and corresponds to the velocity measured. Due to the damping of this shock wave it finally reduces to an elastic wave, and the corresponding velocity of propagation is indeed measured for the greatest thicknesses of lead used.

§ 2. THE VELOCITIES OF WAVES IN METALS

Before proceeding, it may be of use to recall the nature of the different wave velocities which may be measured in metals.

For elastic disturbances in which the displacement is wholly in the direction of propagation of the waves, the waves are said to be "plane", and the velocity is given by

$$V_1 = \sqrt{\left\{\frac{3K}{\rho}.\frac{(1-\nu)}{(1+\nu)}\right\}},$$

where K is the bulk modulus of compression, ρ the density, and ν is Poisson's ratio. This is also the general "velocity of sound" for waves of dilation, involving neither shear nor rotation, and is not be to confused with the "velocity of sound" in experiments on rods, in which lateral displacement is freely allowed, and in which, on the assumption that a purely longitudinal stress may be propagated along a rod without accompanying transverse stresses, the velocity of sound is $V_2 = \sqrt{(E/\rho)}$, where E is Young's modulus. The velocity V_2 and the velocity of transverse waves $V_3 = \sqrt{(G/\rho)}$, where G is the modulus of rigidity, are not relevant to the problems under discussion in the present work, in which the conditions are such that we avoid the necessity of considering the transverse displacements (see § 3).

When the stress is at the yield point of the medium, "plastic waves" similar in nature to the plane waves, are propagated with velocity $V_4 = \sqrt{(K/\rho)}$. Since Poisson's ratio cannot exceed $\frac{1}{2}$, a result which is deducible from the relations between the elastic constants (see Southwell 1941), it follows at once that $V_4 \leqslant V_1$, and we may say that for all metals the plastic waves are initially slower than the plane elastic waves. However, for high amplitudes of pressure in the applied pulse, the plastic waves are capable of achieving very high velocities, if, as strain increases, the rate of change of stress with strain also increases; in other words, if the effective modulus of compression increases with strain. The higher stresses are propagated with ever increasing velocities, and so lead to a shock wave, whose velocity eventually exceeds the velocity of the elastic waves.

The waves considered above (other than shock waves) have been calculated to have velocities given (approximately) in table 1.

Table 1

	$\frac{1}{K}$ (dynes/cm^2)	ρ (gms./c.c.)	ν	V_1 (metres/sec.)	V_2 (metres/sec.)	V_3 (metres/sec.)	V_4 (metres/sec.)
Iron	$0{\cdot}61 \times 10^{-12}$	7·8	0·28	5950	5000–5500	3200	4600
Lead	$2{\cdot}20 \times 10^{-12}$	11·4	0·45	2150	1200	700	2000

The difference between the velocity of plane elastic waves and of elastic waves in rods is clearly seen in the above table, particularly for lead, and is due to the large value of Poisson's ratio for this metal. For average values of ν (say $\nu = 0{\cdot}3$) the difference between V_1 and V_2 is not sufficiently marked to make the velocities experimentally distinguishable.

§3. THE STRESS SYSTEM

When an impulsive force such as an explosion occurs over a surface, the condition immediately beneath the surface is virtually one in which lateral strain is absent. This is because it is possible to choose a small thickness of the block such that the time in which the pulse from the explosion traverses this thickness is not sufficient to allow the lateral displacement which is necessary to produce lateral strain to take place across the much greater area of the impulse. The limiting depth for which the condition of no lateral strain may be assumed depends, therefore, upon the area over which the explosion occurs; and we see that we may make this assumption for a block of which the thickness is small compared with the area of the explosion. Also, if, as in our experiments, the explosion occurs over an area which is less than the whole area of block surface, the consideration above assures that there will be no interference near the axis from waves which have been reflected from the lateral boundary of the block. Under these conditions the principal axes of strain are along the axis of the block and in the plane normal to it. Denoting the principal strains by $\epsilon_1, \epsilon_2, \epsilon_3$, and the corresponding principal stresses by $\sigma_1, \sigma_2, \sigma_3$, the suffix unity referring to the axial component, we have $\epsilon_2 = \epsilon_3 = 0$, $\sigma_2 = \sigma_3$, and $\sigma_1 = p$, the uniform pressure. The components σ_2, σ_3 are set up as a direct result of the effective prevention of lateral strain by inertia, in contrast with the case of a thin bar under compression, where these components vanish on account of the freedom for expansion.

R. von Mises has given a condition for the onset of plastic flow (see, for example, Nadai 1931), viz.:—

$$(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2=2Y^2,$$

where Y is the yield stress in simple tension.

The left hand side of this expression is proportional to the elastic energy per unit volume used in changing the shape, i.e. the potential energy of distortion, as opposed to that causing volume change under elastic strains. The condition therefore states that plastic flow begins when the potential energy of distortion reaches a limiting value, and that this energy remains constant *during* the plastic flow.

Putting $\sigma_2 = \sigma_3$, we see that for the conditions under discussion, the equation reduces to

$$|\sigma_1-\sigma_2| = Y.$$

The volume change in the metal is due to that part of the stress system which corresponds to a hydrostatic pressure σ equal to the mean of the applied stresses, i.e. $\sigma = \frac{1}{3}(\sigma_1+\sigma_2+\sigma_3) = \sigma_1 - \frac{2}{3}Y$; it is clear that if the pressure applied is so large that Y may be neglected in comparison, the change of density may be put equal to that due to a hydrostatic compression of magnitude equal to the applied pressure. Shear distortion and plastic flow do not contribute to this effect.

Let the volume of the metal under applied pressure p, be changed from V_0 to V, the decrement being denoted by $-\Delta V$.

Let us also write

$$p = \alpha\left\{\exp\left(\beta\frac{\sqrt[3]{V_0}-\sqrt[3]{V}}{\sqrt[3]{V_0}}\right)-1\right\}\Big/\left(\frac{V}{V_0}\right)^{2/3} \qquad \ldots\ldots(1)$$

This tentative law is suggested by quantum mechanical considerations, since it requires the pressure to increase exponentially with the change in the interatomic distance, and has been constructed so that $V=V_0$ when $p=0$, and $V\to 0$ as $p\to\infty$. α, β are two parameters, by means of which an extrapolation to high pressures can be made from the results determined in experiments on volume changes due to hydrostatic compression at lower pressures. The extrapolation involved is from 30000 atmospheres for steel, 12000 atmospheres for lead to (approx.) 280000 atmospheres at the surface of a target.

To determine the values of α, β we expand (1), substituting a power series in p for $-\Delta V/V_0$. To the second degree in p we obtain:—

$$-\frac{\Delta V}{V_0}=\frac{3}{\alpha\beta}p-\frac{3}{2\alpha^2\beta^2}(\beta+6)p^2. \qquad \ldots\ldots(2)$$

This may be compared with the empirical formulae deduced by Bridgman from experimental results on many different metals, and written in the form

$$-\frac{\Delta V}{V_0}=Ap-Bp^2.$$

In what follows, the variability of A, B with temperature is not taken into account.

For iron (Bridgman 1940), we use $A=5{\cdot}94\times10^{-13}$, $B=0{\cdot}60\times10^{-24}$ expressed in c.g.s. units. These figures lead to the values

$$\alpha=120{\cdot}16\times10^{10}\text{ dynes/sq. cm.},\quad \beta=4{\cdot}203.$$

For lead (Bridgman 1931), we use $A=24{\cdot}50\times10^{-13}$, $B=18{\cdot}16\times10^{-24}$ expressed in c.g.s. units. from which we obtain

$$\alpha=10{\cdot}08\times10^{10}\text{ dynes/sq. cm.},\quad \beta=12{\cdot}152.$$

It may be remarked in passing that the values A, B used above lead to values of the thermal expansion coefficient which are in good agreement with those predicted theoretically by Mott and Jones (1936); the original value of B for iron, quoted by Bridgman (1931), appeared to be too large, and has, in fact, been corrected to that used in the above calculations, as the result of experiments at higher pressures than had previously been attainable (see Bridgman 1940).

§4. THE SHOCK-WAVE VELOCITY

We now calculate the velocity at which a shock wave advances into a metal, using the relationship between pressure and density which we have already postulated in equation (1).

Suppose the shock wave moves with a velocity U into a metal at rest. Let the velocity of the metal after passage of the wave be u in the same direction, and let its density be changed from ρ_0 to ρ. Let the pressure jump be p.

Then the equation of continuity is

$$\rho(U-u)=\rho_0U, \qquad \ldots\ldots(3)$$

and the momentum equation is

$$p=\rho_0Uu, \qquad \ldots\ldots(4)$$

from which we obtain

$$U^2=\frac{\rho}{\rho_0}\cdot\frac{p}{\rho-\rho_0}. \qquad \ldots\ldots(5)$$

Equation (1) supplies the necessary additional information to correlate the density change with the pressure.

Corresponding to a given pressure p_0 in the detonation wave front, there will be a pressure p at the surface of the target, determined by the necessity for the pressure and mass-velocity of the metal to be equal to the pressure and mass-velocity in the explosion gases, after the propagation, from the interface, of a shock wave forwards into the metal, and backwards into the gases. The gas relationships for various high explosives have been worked out by Dr. H. Jones, and are shortly to be published. Dr. A. F. Devonshire kindly supplied us with tables based upon the relationships for the explosive used in our experiments; by assuming a pressure in the detonation wave front of 17×10^{10} dynes/sq. cm., i.e. 1100 tons/sq. in., we were able to deduce that:—

at the surface of a steel target, $p = 28{\cdot}3 \times 10^{10}$ dynes/sq. cm.

at the surface of a lead target, $p = 27{\cdot}1 \times 10^{10}$ dynes/sq. cm.

From these we obtain:

for steel $U = 5240$ m./sec. and $u = 690$ m./sec.;

for lead $U = 3020$ m./sec. and $u = 790$ m./sec.

These examples are interesting since they indicate two distinct possibilities. The velocity of the shock wave in steel is seen to be less than the velocity of the plane elastic waves, which therefore travel in advance of the shock wave; in lead, however, the shock wave has an initial velocity well in excess of the velocity of the elastic waves, and is therefore the fastest pulse in this metal.

§5. EXPERIMENTAL METHODS

To examine experimentally the velocities of the fastest pulses, charges of explosive were detonated in contact with bars of steel and of lead. We see from what has been written above (§2 and §4), that steel and lead may be expected to provide excellent examples of two different phenomena. For steel the initial shock wave velocity has been calculated to be less than that of the fastest elastic waves, while for lead it was found to move at almost $1\frac{1}{2}$ times the velocity of these plane waves, under the pressures set up by the explosive.

The charges used were made up of 45 gm. explosive, confined in copper tubing 1/10 in. wall thickness and $1\frac{1}{4}$ in. internal diameter. The front of each charge was flat and unconfined, the rear being in the form of an 80° cone tamped with 15 g. plasticene and initiated by a No. 8 Briska detonator. The targets were 3 in. in diameter, and were of various lengths. The steel was cut from a bar of about 0·3% carbon and about 35 tons/sq. in. ultimate stress. The lead bars were carefully cast and then cut to the required lengths.

A diagram of the arrangement employed for the velocity measurements is given in figure 1. The front of the charge is in contact and central with one face of the target. Central with the other face is a 3/16 in. polished steel ball soldered to the end of a screw which passes through a wooden support. The screw was turned until electrical contact was just not made between the ball and the target. This meant an air gap of a fraction of a thousandth of an inch between ball and face.

Through the copper tubing and near to the front of the charge is a hole about 0·1 in. diameter. In very close proximity to and central with this hole is an ionization head which consists merely of two enamelled copper wires 26 s.w.g. running side by side, separated from each other by their insulation, but with their

ends bare. Thus, when the ionized gases reach the hole in the tube, the "head" passes a current, and in this way, subject to an error to be discussed later, we are able to record the arrival of the detonation wave at the face of the target. A pulse of compression is propagated through the target, setting the metal in motion as it progresses. The approximate instant it reaches the rear of the target is recorded by the movement of the rear face, making contact with the steel ball.

The apparatus used to register the time difference between head and ball events was a Baird microsecond chronometer.

§6. ERRORS INVOLVED IN THE EXPERIMENTS: CORRECTIONS APPLIED

(i) The input circuits of the contour were first adjusted to trip at the same amplitude (6 v.) of voltage pulse. To overcome tripping instability caused by the long leads to the bomb chamber, a voltage of 66 v. was applied, attenuated by a 10 : 1 (5000, 500-Ω resistors) potentiometer. Low capacity and widely spaced event leads diminished distortion and lag of the pulses (see figure 1).

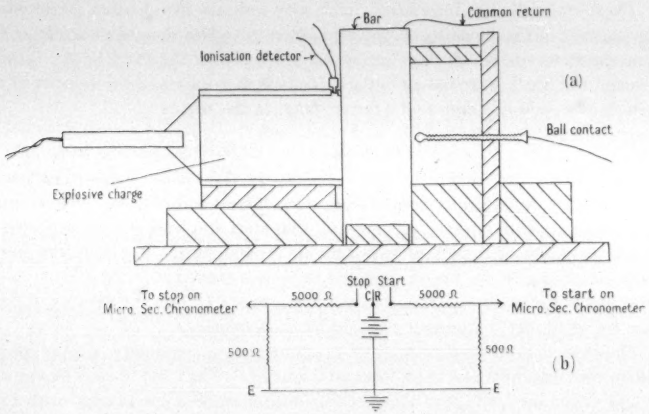


Figure 1.

(ii) The difference of pulse delays in the two leads was measured by transmitting the same pulse down both leads and measuring the delay between their arrivals at the other ends. This time interval was of the order of 0·1 μ sec., the "stop" event lead being the quicker, implying that time readings with these leads were 0·1 μ sec. too low.

(iii) The delay between the impact of detonation wave on the steel surface and the emergence of ionized gases through the hole on the periphery of the charge (normal start event) was determined by placing another ionization detector at the end of the charge (see figure 1). The delay between the two detectors was 0·4 μ sec. which added to (ii) gave a total error of 0·5 μ sec.

(iv) The time-lag involved in the ball event at the end of the metal specimen (figure 1) is determined mainly by the distance between this surface and the ball.

This lag was determined experimentally by finding the delay between two balls at different known distances from the back surface. It was concluded that the maximum error derivable from this source is not greater than 10^{-7} sec.

These errors were taken into account and it is considered that the values of the times obtained are correct to $\pm 1\%$.

After making the corrections outlined above, the times of propagation of the fastest pulse in various lengths of lead and of steel have been plotted against the lengths. The gradient of a curve at a given length measures the velocity of the fastest pulse at an equal distance from the surface of the target (figures 2 and 3).

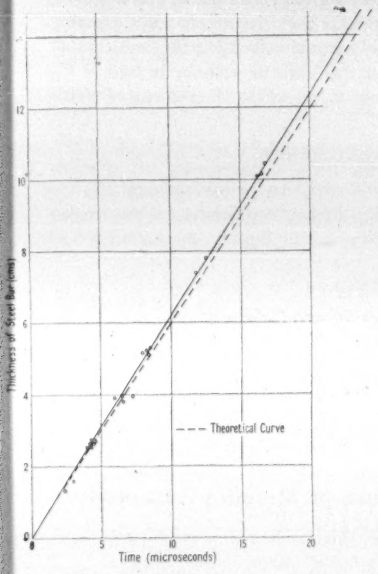


Figure 2.

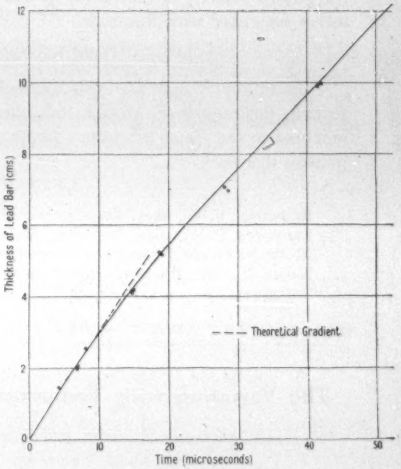


Figure 3.

§7. DISCUSSION OF RESULTS

The experimental values obtained for the times for the fastest wave to traverse the bars of differing lengths are plotted in figures 2 and 3. Figure 2 gives the graph for steel, and it is seen from the experimental curve that there is no appreciable decay in velocity for lengths up to at least 12 in. Since the theoretical value of the shock wave velocity is below that given by V_1 (the general velocity of elastic waves), the theoretical curve has been drawn with gradient corresponding to V_1. The agreement with the experimental curve is very good, so that there can be little doubt that the velocity measured is that of the elastic waves, V_1.

A different picture is presented for the bars of lead. In figure 3 the curve through the experimental points has a pronounced curvature near the origin, becoming straight as the thickness of the bar is increased. The shock wave

velocity, determined theoretically is approximately 3000 m./sec., well in excess of the velocity of elastic waves. A dashed line through the origin represents the gradient associated with the former velocity, and it is seen to lie along the experimental curve at the origin. Since, by extrapolation of the curve to zero length of bar, we obtain the undamped velocity of the shock wave, we should, in fact, expect the value at the origin to correspond to that calculated. For the longest bars of lead used, the gradient of the experimental curve leads to a velocity in close agreement with that of plane elastic waves, viz. 2100 m./sec.

We conclude then, that the pressure set up by the detonation in steel is not sufficient to propagate a shock wave at a velocity greater than that of elastic waves; in lead the shock wave is propagated much faster than elastic waves but damping occurs very rapidly with increasing length of bar, the velocity finally being that of elastic waves. It is to be emphasized that this ultimate velocity in lead is V_1, the general velocity of elastic waves, and not $V_2=\sqrt{(E/\rho)}$, the velocity of elastic waves associated with thin rods.

ACKNOWLEDGMENTS

The authors wish to acknowledge the kind advice and encouragement given to them by Professor N. F. Mott, F.R.S., during these investigations. They are also indebted to the Chief Scientific Officer, Ministry of Supply, for permission to publish this work.

REFERENCES

BRIDGMAN, P. W., 1931, *The Physics of High Pressures* (London).
BRIDGMAN, P. W., 1940, *Phys. Rev.*, **57**, 235, 273 and 342.
MOTT, N. F., and JONES, H., 1936, *Properties of Metals and Alloys* (Oxford).
NADAI, A., 1931, *Plasticity* (New York and London).
SOUTHWELL, R., 1941, *Theory of Elasticity*, 2nd edition, p. 124 (Oxford).

The Variation with Temperature of Metallic Reflectivity

BY ROBERT WEIL
South-West Essex Technical College

MS. received 22 *May* 1947

ABSTRACT. It is shown that, assuming free electrons alone to be responsible for the mechanism of metallic reflection, there is a wavelength at which the temperature variation of reflectivity is zero. Since this wavelength is directly proportional to the time of relaxation of the electrons it is to be concluded that it varies with the temperature. The effect of the bound electrons is also considered.

§1. INTRODUCTION

IN a recent communication (1947) it was pointed out by Price that the temperature variation of the emissivities of several metals is such as to suggest critical wavelengths, specific to each metal, for which the temperature coefficient of emissivity is zero. These wavelengths were called X-points. Several explanations were advanced but it would appear to the present writer

that these X-points can be predicted if dealt with from the point of view of reflectivity, and that, as a result of such an investigation, Price's conclusions would have to be scrutinized more closely.

§ 2. THEORY

(a) *Free electrons*

While the temperature coefficient of emissivity is defined by

$$\beta = \frac{1}{E} \cdot \frac{dE}{dT}, \qquad \ldots\ldots(1)$$

it is desirable to consider the gradient dE/dT only, because this is simply related to the corresponding reflectivity gradient by the expression

$$\frac{dE}{dT} = -\frac{dR}{dT}. \qquad \ldots\ldots(2)$$

Obviously β will be zero when dE/dT is zero, and, therefore, it will be in order to consider the wavelength at which dR/dT is zero: this, then, will be an X-point.

Electro-magnetic theory leads to the following expressions connecting the reflectivity R, the refractive index n, the extinction coefficient k, the conductivity σ, and the dielectric constant ϵ with the frequency of the radiation ν:

$$nk = \frac{\sigma}{\nu}, \qquad \ldots\ldots(3)$$

$$n^2 - k^2 = \epsilon. \qquad \ldots\ldots(4)$$

$$R = 1 - \frac{4n}{(n+1)^2 + k^2}. \qquad \ldots\ldots(5)$$

It is assumed that the permeability μ of the metal to which these expressions apply is equal to unity.

On differentiating (5) with respect to the temperature T, we obtain

$$\frac{dR}{dT} = \frac{4}{D^2}\left[(\epsilon - 1)\frac{dn}{dT} + 2\frac{\sigma}{\nu}\frac{dk}{dT}\right], \qquad \ldots\ldots(6)$$

where $$D = (n+1)^2 + k^2.$$

Multiplying and dividing the above expression by n:

$$\frac{dR}{dT} = \frac{4}{nD^2}\left[(\epsilon - 1)n\frac{dn}{dT} + 2n^2 k\frac{dk}{dT}\right]. \qquad \ldots\ldots(7)$$

The following should be borne in mind in the continuation of the calculation. According to Price, the region connected with the X-point is a transition region, i.e. it cannot be assumed that the wavelength under consideration is either "very long" or "very short". It is, therefore, imperative not to carry out any premature approximations, however complicated the calculation may become. The treatment below is based on the assumption that only free electrons are involved in the process of reflection: the effect of the bound electrons is to be considered later.

It has been shown by Zener (1933) that the equations (3) and (4) can be made to account for experimental results in the infra-red and visible parts of the spectrum,

if, in addition to the direct-current conductivity σ_0, the time of relaxation τ is taken into consideration. He obtained these expressions:

$$n^2-k^2=\epsilon=1-\frac{4\pi N\boldsymbol{e}^2}{m}\cdot\frac{1}{\omega^2+1/\tau^2}, \qquad \ldots\ldots(8)$$

$$2nk=2\frac{\sigma(\nu)}{\nu}=\frac{4\pi N\boldsymbol{e}^2}{m}\cdot\frac{1}{\omega\tau}\cdot\frac{1}{\omega^2+1/\tau^2}. \qquad \ldots\ldots(9)$$

Here $$\tau=\frac{\sigma_0 m}{N\boldsymbol{e}^2},$$

where N is the number of free electrons per c.c., and $\boldsymbol{e}$ and m represent the electronic charge and mass respectively. $\sigma(\nu)$ is the conductivity corresponding to the frequency $\nu(=\omega/2\pi)$. It will simplify the presentation of the formulae if we put $$\frac{1}{\tau}=\nu_\tau \quad \text{and} \quad \frac{4\pi N\boldsymbol{e}^2}{m}=\alpha.$$

From (3) and (4) it follows

$$\left.\begin{aligned} 2n^2&=\sqrt{\epsilon^2+4\frac{\sigma^2}{\nu^2}}+\epsilon,\\ 2k^2&=\sqrt{\epsilon^2+4\frac{\sigma^2}{\nu^2}}-\epsilon. \end{aligned}\right\} \qquad \ldots\ldots(10)$$

Substituting (8) and (9) in (10)

$$2\frac{n^2}{k^2}=\sqrt{1+\frac{a^2-2a\omega^2}{\omega^2(\omega^2+\nu_\tau^2)}}\pm 1\mp\frac{a}{\omega^2+\nu_\tau^2}. \qquad \ldots\ldots(11)$$

Assuming that N and thus a are constants,

$$\left.\begin{aligned} n\frac{dn}{dT}&=\left[\frac{2\omega^2-a}{4\omega^2\sqrt{1+\dfrac{a^2-2a\omega^2}{\omega^2(\omega^2+\nu_\tau)}}}+1\right]\frac{a\nu_\tau}{(\omega^2+\nu_\tau^2)^2}\cdot\frac{d\nu_\tau}{dT},\\ \text{and}\qquad k\frac{dk}{dT}&=\left[\frac{2\omega^2-a}{4\omega^2\sqrt{1+\dfrac{a^2-2a\omega^2}{\omega^2(\omega^2+\nu_\tau^2)}}}-1\right]\frac{a\nu_\tau}{(\omega^2+\nu_\tau^2)^2}\cdot\frac{d\nu_\tau}{dT}. \end{aligned}\right\} \qquad \ldots\ldots(12)$$

Substituting (8), (11), and (12) in (7), and considering only the expression which can make dR/dT equal to zero, then, if $\surd$ represents the square root in (12),

$$\frac{dR}{dT}\propto\left\{\left(-\frac{a}{\omega^2+\nu^2}\right)\left[\frac{2\omega^2-a}{4\omega^2\surd}+1\right]+\left[1-\frac{a}{\omega^2+\nu_\tau^2}+\surd\right]\left[\frac{2\omega^2-a}{4\omega^2\surd}-1\right]\right\}. \qquad \ldots\ldots(13)$$

At the X-point,

$$\left[1-\frac{a}{\omega^2+\nu_\tau^2}+\surd\right]\left[\frac{2\omega^2-a}{4\omega^2\surd}-1\right]=\frac{a}{\omega^2+\nu_\tau^2}\left[\frac{2\omega^2-a}{4\omega^2\surd}+1\right].$$

A simple transformation leads to

$$\sqrt{1-\frac{2\omega^2-a}{\omega^2+\nu^2}\cdot\frac{a}{\omega^2}}=\frac{2a}{2\omega^2+a}\left(\frac{2\omega^2-a}{\omega^2+\nu_\tau^2}\right)-1. \qquad \ldots\ldots(14)$$

Writing $\omega^2 = x$, $\nu_\tau^2 = y$, $(2\omega^2 - a)/(\omega^2 + \nu_\tau^2) = z$, and squaring (14)

$$\frac{4az}{(2x+a)^2} - \frac{4}{(2x+a)} + \frac{1}{x} = 0.$$

$$\therefore\ 4ax\frac{2x-a}{x+y} - 4x^2 + a^2 = 0,$$

$$x^3 - (2a-y)x^2 + \frac{3a^2}{4}x - \frac{a^2 y}{4} = 0. \qquad \ldots\ldots(15)$$

When solved, this equation gives rise to a complicated expression. But a few interesting results can be obtained without solving it.

By differentiating (15) with respect to x it is found that a maximum and a minimum occur at imaginary points. Further, when $x = -\infty$ the above expression is equal to $-\infty$; when $x = 0$, the expression equals $-a^2y/4$; and when $x = \infty$ the expression is equal to $+\infty$. It follows that the above function cuts the x-axis only at one point, giving a positive value for x. Since x represents the square of ω, and the negative value of $\omega = \pm\sqrt{x}$ is rejected, it is seen that there is only one such X-point due to the free electrons.

An inspection of (13) indicates that when ω is very small as compared with ν_τ and a, dR/dT will have a negative sign, as would be expected from the Hagen-Rubens relation.

It follows from (12) that the X-point is a relaxation phenomenon since (15) will be true also for $dR/d\nu = 0$.

On substituting numerical values it is found that the first two terms are negligible as compared with the latter two (15). Then, since

$$\sqrt{y} = \nu_\tau = 1/\tau = 2\pi c/\lambda_\tau$$

$$\lambda_x = \sqrt{3}\,\lambda_\tau. \qquad \ldots\ldots(16)$$

This is in disagreement with figure 2*a* of Price's paper: there it would appear that λ_x was smaller than λ_τ. Moreover, since λ_τ is a function of the temperature, λ_x must be one likewise. The temperature coefficient of λ_τ is given by that of the conductivity (Mott and Jones 1936). Therefore λ_x will be a constant for large temperature ranges only if that coefficient is negligibly small.

It will be clear from figure 1 how Price arrived at the belief that there existed an X-point whose value would be independent of the temperature. The group of curves gives a picture of the above analysis. In his work Price obtained the X-point by averaging emissivity measurements at high and reflectivity measurements at low temperatures respectively. From figure 1 it is seen that the reflectivity curves for different temperatures T_1 and T_4 cut each other at one point, but those for T_1 and T_2 do so at another, etc. It is significant that Hurst (1933) and Reid (1941), who both obtained X-points in connection with nickel, should obtain different values for this point which is supposedly characteristic of each metal, the difference being explicable in terms of the different temperature ranges in which the two workers carried out their experiments. It is only fair to admit the above theory has not tallied with all the available experimental results but to this point we shall return below. However, in spite of the scantiness of the data, it appears that bad conductors exhibit shorter wavelengths λ_x than good ones as would be expected on the basis of (16).

Theoretically speaking, the measurement of the reflectivity or emissivity, in a direction perpendicular to the surface in question, at different temperatures, and in this transition region, can provide values for τ and a ($=4\pi Ne^2/m$), which have, so far, been obtainable from catoptric measurements only. It should be pointed out that it is unlikely that the accuracy of an experimental arrangement based on this principle could emulate that of a catoptric method.

(b) *Bound electrons*

Owing to the larger number of variables in the case of bound electrons it does not seem possible to make any reasonable predictions about the temperature effect.

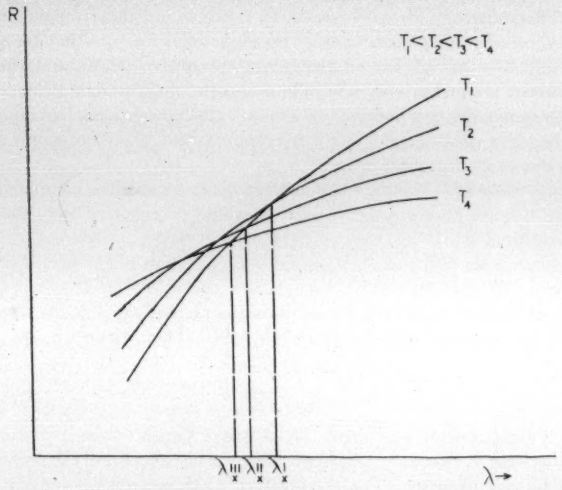


Figure 1

These variables are: the wavelength λ_0 of the absorption band, the band-width λ', as defined by Wood (1936), and the number of the bound electrons taking part in the vibration. As might be expected the (highly complex) expression shows that R will be affected differently according as the wavelength λ is greater or less than λ_0. Not even far-reaching approximations yield any more detailed information.

In this connection it is doubtful whether figure 3 of Price's paper is justifiable. While the absorption curves for T_1 and T_2 very likely cut each other as indicated there, it is difficult to see why the interpolated curve should cut them in the same point. Sélincourt's results for silver (1925) show a region of small temperature variation, but it is found to be on the short-wave side of the absorption band. The presence of bound electrons with a natural wavelength λ_0 in the neighbourhood of λ_x as due to free electrons only, will modify the value of the latter.

§ 3. CONCLUSION

We doubt whether an X-point as previously defined has any theoretical significance; until a more satisfactory theory is propounded it is to be assumed

that, whenever it occurs over a large range of temperatures, it is due to a fortuitous combination of the effects of free and bound electrons. However, while λ_x as defined in the present paper depends on the temperature, its possible significance in connection with optical pyrometry should not be overlooked, and will be dealt with on another occasion.

§4. ACKNOWLEDGMENT

The writer would like to thank Dr. H. Lowery, F.Inst.P., for his interest in the above problem.

REFERENCES

HURST, 1933, *Proc. Roy. Soc.*, **142,** 466.
MOTT and JONES, 1926, *Theory of the Properties of Metals and Alloys*, p. 119 (Oxford : Clarendon Press).
PRICE, 1947, *Proc. Phys. Soc.*, **59,** 131.
REID, 1941, *Phys. Rev.*, **60,** 161.
SÉLINCOURT, 1925, *Proc. Roy. Soc.*, **107,** 247.
WOOD, 1936, *Physical Optics*, p. 487 (New York : The Macmillan Company).
ZENER, 1933, *Nature, Lond.*, **132,** 968.

Work Function and Energy Levels in Insulators

BY D. A. WRIGHT
(Communication from the Research Staff of the M.O.Valve Co. Ltd. at the G.E.C. Research Laboratories, Wembley, England)

MS. received March 1947; *in revised form* 2 *May* 1947

ABSTRACT. An estimate is made for several insulators of the energies of the highest filled energy band and of the empty conduction band. For BaO, SrO, CaO, MgO and BeO, the bottom of the conduction band is near the zero level. It is considerably lower in AgBr, ZnO and ZnS. The bearing of the results on thermionic emission, secondary emission and photoconductivity is briefly discussed, with special reference to BaO and SrO.

§1. WORK FUNCTIONS AND ENERGY LEVELS OF INSULATORS

CALCULATIONS concerning the position of the conduction band and of energy levels which determine the ultra-violet absorption bands were made by Mott (1938) for the case of some of the alkali halides. In these calculations it was necessary to use estimates of polarisation energies as made in a previous paper (Mott and Littleton 1938). The theory is also discussed in *Electronic Processes in Ionic Crystals* (Mott and Gurney 1940) to which frequent reference will be made. In the present paper considerations of this type are used to estimate the probable maximum value of the work function for electron insertion for some silver halides and divalent oxides. Here it is necessary to consider band widths and the effect of the selection rules.

§2. METHOD FOR ALKALI HALIDES

We first outline Mott's procedure in the case of the alkali halides.

The potential energy of an electron at a great distance from the crystal is indicated by the zero level O in figure 1. A is the bottom of the conduction band, so that $-\chi$ is the work necessary to introduce an extra electron into the crystal, placing it on a positive ion. C is the top of the first fully occupied band, ϕ is therefore the work to extract an electron from this band, which is usually associated with the negative ion. B is the first excited level corresponding to the formation of an exciton when an electron is removed from the full band, but remains in the field of the positive hole. If an electron is removed from this field the case corresponds with ionisation, the electron enters the conduction band, and the necessary energy is $\phi-\chi$. This energy should correspond with $\boldsymbol{h}\nu_s$ where ν_s is the frequency at the series limit of the absorption bands. The energy θ corresponds with $\boldsymbol{h}\nu_1$ where ν_1 is the frequency of the longest wavelength absorption band. In this treatment the width of the energy level bands was neglected, so that forbidden transitions were not considered.

The procedure was to suppose that an ion is removed from the crystal, is converted to an atom, and the atom is replaced. This either adds an electron to a positive ion or removes one from a negative ion. The method is fully discussed in Mott and Gurney (1940), pages 71, 80 and 97.

The results are

$$-\chi = W_L - I + \omega_2 + p, \qquad \ldots\ldots(1)$$

$$\phi = W_L + E + \omega, \qquad \ldots\ldots(2)$$

$$\theta = (2\alpha - 1)\frac{\boldsymbol{e}^2}{r} + E - I + \omega_1 + p'. \qquad \ldots\ldots(3)$$

Here W_L is the lattice energy per ion pair $= \alpha\boldsymbol{e}^2/r - R$,

α is the Madelung constant for the crystal, r the interionic distance, $\boldsymbol{e}$ the electronic charge and R the energy of repulsion between an ion and its six nearest neighbours.

I is the ionisation potential of the alkali atom.

E is the electron affinity of the halogen atom.

The terms ω, ω_1 and ω_2 are polarisation energies, which will be discussed below. p is a term to allow for the fact that energy is required to insert an atom in the lattice compared with the smaller positive ion which was extracted. Such a term is omitted in (2) since an atom replaces a larger negative ion. In (3), p' will be similar to p, but not identical with it, since in the case (3) of electron transfer, a neighbouring ion has been converted to an atom.

In these equations all the terms can be calculated except p and p'. For the alkali halides, experimental values are available for $\boldsymbol{h}\nu_1 = \theta$ and for $\boldsymbol{h}\nu_s = \phi - \chi$. Taking experimental values of $(\phi - \chi)$ and calculated values of ϕ, Mott and Gurney (p. 97) deduced values of χ. They also showed that the calculated values of $(\theta - p')$ were about 2 ev. greater than the observed values of $\boldsymbol{h}\nu_1$ (see page 99 of this book).

The lattice energy W_L should, however, be used in all three cycles, whereas Mott employed it only in the first two cases. Use of the energy $W_L = \alpha\boldsymbol{e}^2/r - R$ in the third cycle would give

$$\theta = (2\alpha - 1)\boldsymbol{e}^2/r - 2R + E - I + \omega_1 + p'. \qquad \ldots\ldots(4)$$

Then in all cases the value of $(\theta - p')$ would be in good agreement with the observed values of $h\nu_1$, indicating that p' is nearly zero.

It is not, however, important to discuss values of p' since the calculation of θ by equations (3) or (4) is not satisfactory. In the excited state the electron which is removed from the full band is not localised on any particular neighbouring ion. Thus the agreement with experiment is probably fortuitous, and calculations of θ will not be relied upon as an essential part of our later argument. It is perhaps preferable to regard the excited electron in the field of the positive hole as a hydrogen-like system in a medium of dielectric constant K_0. Compared with the hydrogen case, force between charges is then multiplied by $1/K_0$, and energy separations are multiplied by $1/K_0^2$. Thus if the separation $(h\nu_e - h\nu_1)$ is known

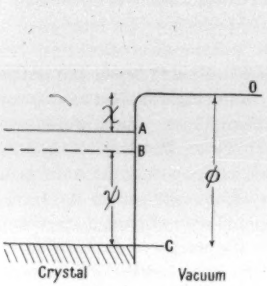


Figure 1. Energy levels in insulator.

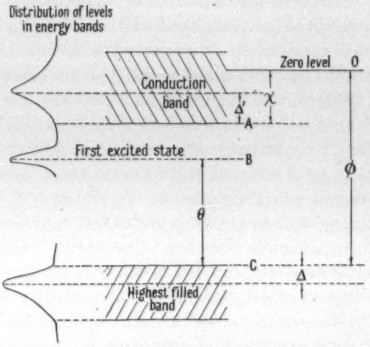


Figure 2. Energy levels in insulator including band width.

A is the bottom of the empty band.
B is the first excitation level.
C is the top of the upper fully occupied band.

for one crystal, it can be estimated for others of known K_0 since it is proportional to $1/K_0^2$. K_0 is here the dielectric constant at infra-red frequencies.

§3. ENERGY OF POLARISATION

The dipole formed by the electron transfer from negative ion to positive in process (3) polarises the surrounding medium, with an energy given by de Boer (1935) as

$$\omega_1 = -2{\cdot}03\,\frac{e^2}{r^4}\,(\alpha_1 + \alpha_2),$$

where α_1 and α_2 are the polarisabilities of the ions.

When an electronic charge is added or subtracted at a lattice site, the surrounding medium is polarised, and the induced dipoles produce a potential V at the site in question. The energy of this polarisation is then $-\frac{1}{2}(eV)$. Mott and Littleton (1938) gave methods for calculating ω and ω_2, which are also described in Mott and Gurney (1940) pages 58 and 59. Our procedure has been to use Mott's approximation of the first order, and to calculate the potentials when an electron is

added at a positive ion site, ω_2, or subtracted at a negative ion site, ω. In all cases the estimation of α_1 and α_2 has been made taking the observed value of K_0 which gives $(\alpha_1+\alpha_2)$ according to the formula

$$\frac{K_0-1}{K_0+2}=\frac{4\pi}{3}\frac{1}{r^3}\frac{\alpha_1+\alpha_2}{2}.$$

Then we take the observed ratio α_1/α_2 for the free ions and determine the values of α_1 and α_2 from α_1/α_2 and $\alpha_1+\alpha_2$. The values found in this way are shown in the tables below, where observed values are those given by Mott and Gurney (1940).

§4. BAND WIDTH

In the considerations so far, band width has been neglected. We shall make no attempt to calculate band widths in the insulators in which we are interested, but shall consider the experimental evidence available, and the effect of observed band widths on calculations of the type discussed above. Now O'Bryan and Skinner (1940) showed that in the alkali halides the width of the highest filled band is from 5 to 19 ev., and in a variety of oxides it is from 14 to 23 ev. Such widths must have a considerable effect on our calculations. However, the shape of the bands must be considered, and we note that in most cases a large part of the width is due to ends where the electron density is very low. The main part of the band is frequently a single sharp peak which may be compared with a broadened spectrum line. In such cases the "line-width", which is the width of the band at half maximum intensity, is considerably smaller than the total width. This is especially true of the alkali halides, and among the oxides of BaO, SrO and CaO, where the line-width is only 2 ev., although the total width is 14 ev. The fact that there is some experimental agreement in the case of the alkali halides makes it doubtful whether the band ends have sufficient density to affect the optical behaviour of the crystal.

We shall suppose here that the calculations leading to equations (1) and (2) above predict the distance below the zero level of the peaks of the energy bands, as indicated in figure 2, and we shall also suppose that the band widths are the widths without the low density ends. Thus for BaO with total width 14 ev., at either side of the band there are ends extending for 4 ev., and we shall take the effective width as 6 ev. The peak is then 4 ev. below the top of the band. It should be noted that apart from the effects of low density of electrons in the top end of the filled band and of low density of levels in the lower end of the conduction band, the transition from the top of the filled band to the bottom of the conduction band is forbidden, and the frequency for the series limit will therefore in any case be greater than that which corresponds to the separation between the band ends. Similarly the transition from the top of the full band to the first excitation level is forbidden. We shall continue to calculate the energy θ from equation (4) for the sake of completeness, and because of the agreement with experiment in the alkali halides; θ should presumably correspond with the separation between the peak of the full band and the peak of the band of excitation levels. However, as pointed out above, the calculation is really unsatisfactory in this case.

We shall in the following refer to figure 2 in our calculations, thus equation (2) determines the depth of the peak of the band below the zero level, and we write ϕ

for the depth of the top of the band, and Δ for the depth of the peak below the top. Thus equation (2) determines $(\phi+\Delta)$, and equation (2) determines $-(\chi-\Delta')$. We shall continue to write (θ) for the energy determined by equation (4).

§ 5. NaCl INCLUDING BAND WIDTH

We first consider the case of NaCl when band width is not neglected. The quantities necessary for the cycle calculations are shown in table 1.

We find

$$\phi+\Delta=10{\cdot}2,$$
$$-(\chi-\Delta')=0{\cdot}3+p,$$
$$\theta=7{\cdot}7+p'.$$

We know that the 3*s* band in the alkali chlorides is narrow and has short ends, thus Δ in this case is about 0·5 ev., hence $\phi=9{\cdot}7$. The excitation level, which is known from experiment to lie 7·7 ev. above C, is therefore 2 ev. below the zero level 0, and the series limit is a little higher than 0. Now we know that pure NaCl is not photo-conducting when irradiated in its first absorption band, thus the bottom of the empty conduction band is considerably higher than the first excitation level, and is therefore not more than about 1 ev. below the zero level. Thus the maximum value of χ is about 1 ev., which is consistant with Mott and Gurney's experimental value of 0·5 ev., page 74. If p is nearly zero, this indicates that Δ' is about 1 ev. Note that when band widths are included the position of the bottom of the conduction band is not necessarily determined by adding to the calculated value of ϕ the observed value of $\boldsymbol{h}\nu_s$, as it was when band widths were neglected. There may now be forbidden transitions with energy lower than $\boldsymbol{h}\nu_s$, so that χ may be greater than zero, and the argument from photo-conductivity becomes necessary.

Table 1

	NaCl	AgBr	AgCl
α	1·75	1·75	1·75
r	2·81	2·88	2·77
$\alpha e^2/r$	8·9	8·6	9·0
K_0	2·33	4·62	4·01
a_1 for halide ion	3·0	3·0	3·0
a_2 for metal ion	0·15	4·1	2·2
ω_1	−1·6	−2·7	−2·7
ω_2	−2·5	−2·9	−3·2
ω	−1·5	−2·8	−2·6
I	5·12	7·54	7·54
E	3·75	3·55	3·75

§ 6. SILVER HALIDES

For AgBr, using the figures in table 1, we find:—

$$\phi+\Delta=8{\cdot}5,$$
$$-(\chi-\Delta')=p-3{\cdot}1,$$
$$\theta=3{\cdot}5+p'.$$

For AgBr, O'Bryan and Skinner found that the width of the main part of the 4*p* band is 4 ev. (1940), and the peak lies 1 ev. below the top of the main part, consequently $\Delta=1$ ev. There is, however, a considerable tail at the top of the band.

The optical absorption evidence (Fesefeldt and Gyulai 1929) shows a rapid increase in absorption beginning near 3000 A., which indicates that the excitation level B lies about 4 ev. above C (figure 2). The high value of K_0 shows that the series limit energy, $h\nu_s$, will only be a little higher than $h\nu_1$, and, in fact, since it is known that photo-conductivity occurs in the tail of the absorption band, the bottom of the conduction band must almost coincide with B. This gives $\chi \simeq 8{\cdot}5 - 1 - 4 = 3{\cdot}5$ ev. This is in good agreement with Mott's value of 3·6 deduced from experiments on the bleaching of emulsions (Mott and Gurney, p. 245). It will be noted that as with the alkali halides, θ agrees roughly with the observed excitation energy if p' is small, and the value of $\chi - \Delta'$ is similar to the value deduced for χ if p is small, indicating that Δ' is probably less than 1 ev. The long optical absorption tail extending into the visible may correspond with transitions from the high-energy tail in the $4p$ band.

For AgCl we find:—

$$\phi + \Delta = 9{\cdot}0,$$
$$-(\chi - \Delta') = p - 2{\cdot}4,$$
$$\theta = 4{\cdot}2 + p'.$$

There is no information concerning the width of the $3s$ band in AgCl, consequently we do not know Δ. As with AgBr, the optical absorption increases rapidly near 4 ev. (Fesefeldt and Gyulai 1929), and Mott here also deduces that χ is about 3·6. This would be consistent on the above argument with $\Delta = 1{\cdot}5$ ev., and with Δ' a little greater than 1 ev.

§7. DIVALENT OXIDES

We suppose that the oxides we shall discuss are in the form of ionic crystals and that the ions are both doubly charged in the normal state.

Now, when the doubly charged positive ion is removed from the crystal, the energy required is $4\alpha e^2/r - R$, and addition of an electron requires the energy $-I_2$, where I_2 is the ionisation potential for the second electron. Replacement in the crystal involves the energy $2\alpha e^2/r - R'$, where R' is the energy of repulsion between the singly charged ion and its nearest neighbours. Here we shall neglect the difference between R and R'. We thus obtain

$$-(\chi - \Delta') = 2\alpha e^2/r - I_2 + \omega_2 + p. \qquad \ldots\ldots(5)$$

Similarly,

$$\phi + \Delta = 2\alpha e^2/r + E_2 + \omega. \qquad \ldots\ldots(6)$$

In the case of the transfer of an electron from a negative to a positive ion, corresponding with excitation, the cycle gives

$$\theta = (4\alpha - 1)e^2/r + E_2 - I_2 + \omega_1 + p'. \qquad \ldots\ldots(7)$$

In these equations, as before, p and p' cannot be satisfactorily calculated. Experimental values of Δ are available, but there is no reliable information about Δ'. Some approximate estimates will be made below. The other quantities are known accurately, except that there may be some uncertainty about the value of E_2, the electron affinity of the singly-charged negative oxygen ion. We take the value deduced by Seitz (1940) of -9 ev., though this might require some modification in

the future. This was obtained by first estimating the affinity of oxygen for two electrons from the results of Sherman (1932) and of Mayer and Maltbie (1932), and then noting that the affinity of neutral oxygen for one electron is +2 ev. (Lozier (1934)).

We are assuming in the above equations that the crystals are completely polar. This was sufficiently true in the case of the alkali halides, and the results of O'Bryan and Skinner (1940) indicate that it is probably a satisfactory assumption for the present purpose in the case of BaO and SrO. They concluded, however, that MgO and BeO are much less completely polar, so that our results will be less accurate in these cases.

The quantities necessary in the calculations are shown in table 2. Here the polarisation energies have been obtained accurately for the cubic crystals, taking Mott's first-order approximation as with the halides. In the case of BeO and ZnO, with a wurtzite lattice, the necessary summations are not available, and we have assumed that for the observed values of the interionic spacing r and the dielectric constant K_0, the terms are similar to those for corresponding values in a cubic lattice.

Table 2

	BaO	SrO	CaO	MgO	BeO	ZnO	ZnS
a	1·75	1·75	1·75	1·75	1·64	1·64	1·64
r	2·77	2·57	2·40	2·10	1·64	1·94	2·33
ae^2/r	9·0	9·8	10·4	11·8	14·4	12·0	10·1
K_0	3·5	3·31	3·28	2·95	2·95	—	5·07
α_1 for negative ion	3·0	2·65	2·9	1·64	0·85	—	3·36
α_2 for metal ion	1·6	0·89	0·6	0·05	0	—	0·12
ω_1	− 2·3	− 2·4	− 2·4	− 2·5	− 3·5	− 2·5*	− 3·5
ω_2	− 2·8	− 2·8	− 2·9	− 3·0	− 2·9	− 3·0*	− 3·7
ω	− 2·4	− 2·3	− 2·0	− 1·7	− 1·6	− 1·5*	− 2·0
I_2	10·0	11·0	11·8	15·0	18·1	17·9	17·9
E_c	− 9·0	− 9·0	− 9·0	− 9·0	− 9·0	− 9·0	− 5·0

* These values are assumed by comparison with the other compounds since K_0 is not known.

§8. BARIUM AND STRONTIUM OXIDES

In table 3 we show the results of the application of equations (5), (6) and (7), the distance Δ from the top of the main part of the filled band to the peak as determined by O'Bryan and Skinner, and hence the value of ϕ.

Table 3

	BaO	SrO
$-(\chi-\Delta')$	$5{\cdot}2+p$	$5{\cdot}8+p$
$\phi+\Delta$	6·6	8·3
θ	$9{\cdot}7+p'$	$11{\cdot}2+p'$
Δ	4·0	5·0
ϕ	2·6	3·3

There is no direct optical information in either case concerning the values of $h\nu_1$ and $h\nu_s$, however, since the crystals do not absorb in the visible, $h\nu_1$ is at least 3 ev. and is probably considerably greater. Thus both the first allowed excitation level and the first ionisation level must lie above the zero line. The bottom of the conduction band will be lower than the first allowed ionisation level, and may in

fact be below the zero line. In the case of BaO we know it cannot be much below the zero line however, since impurity levels can be added to BaO making it a semi-conductor. When excess Ba is present, the crystal becomes semi-conducting with an activation energy varying from 0·5 to 1 ev. according to the experimental conditions. The distance below A (figure 2) of the impurity levels may therefore be as large as 2 ev.; thus C must lie at least 2 ev. below A. Since ϕ is 2·6, we deduce that the maximum value of χ is 0·6 ev. It may of course be less, and will be negative if the bottom of the conduction band is above the zero level.

The experimental results having a bearing on these conclusions will be more fully discussed in the following paper. We may note here, however, that Nishibori and Kawamura (1940) found for both SrO and the equimolecular double oxide (BaSr)O, a value of χ of 0·3 ev., and that they concluded that these values of χ were typical of the oxides themselves, and were not affected by conditions at the outer surface of the oxides, e.g. by the presence of adsorbed Ba. It was primarily in order to find evidence concerning this conclusion that the present calculations were carried out. It will be noted that if we take this experimental value for χ and assume p to be small, as in the case of the halides, we deduce that Δ' is 5·5 ev. for BaO and 6·1 ev. for SrO. These are the depths of the bottom of the conduction levels below their peak. It can also be deduced from the results of Nishibori and Kawamura that the photoelectric work function of the unactivated oxides lies between 2·8 and 3·7 ev. for SrO, and between 2·0 and 2·7 ev. for (BaSr)O. These correspond well with our values for ϕ, as they should if we assume there is no free impurity barium in the unactivated crystal in these experiments.

§9. OTHER DIVALENT COMPOUNDS

In table 4 we show the results on some further oxides and on zinc sulphide, where following Seitz (1940) we take the value of E_2 as -5 ev.

Table 4

	CaO	MgO	BeO	ZnO	ZnS
$-(\chi-\Delta')$	$6{\cdot}1+p$	$5{\cdot}6+p$	$8{\cdot}2+p$	$3{\cdot}0+p$	$-\ 1{\cdot}3+p$
$\phi+\Delta$	9·8	12·9	18·2	13·5	13·2
θ	$12{\cdot}5+p'$	$14{\cdot}0+p'$	$18{\cdot}2+p'$	$11{\cdot}5+p'$	$8{\cdot}3+p'$
Δ	5·0	7·0	7·0	4·0	3·0
ϕ	4·8	5·9	11·2	9·5	10·2

There is little experimental evidence concerning these compounds. For MgO and BeO we have the indirect information obtained by O'Bryan and Skinner, and discussed in Mott and Gurney (1940), pages 77 and 78, which indicates that $h\nu_1$, lies between 11 and 15 ev. for MgO, and is near 14 ev. for BeO. Here again, therefore, excitation raises the electron above the zero level. If we identify θ with $h\nu_1$, we deduce that p' in these cases lies between -2 and -4 ev. We can obtain approximate values of Δ' in the case of BeO and MgO by Seitz's method (1938). In the case of ZnS, Seitz compared the broadening of the 4*s* level of Zn^+ with the broadening of the comparable 4*s* level in Cu, which was studied by Fuchs, and deduced that the depression corresponding with our Δ' was 4 e v. Similarly we can compare the 2*s* levels of Be^+ in BeO with those of metallic Li, and the 3*s* levels of Mg^+ with those of metallic Na, both of which metals have

been studied by O'Bryan and Skinner (1934). Since the separation between the metal ions in the oxides is smaller than between Li and Na atoms in the metal, we conclude that approximately for Be^+, $\Delta' = 5$ ev. and for Mg^+, $\Delta' = 4$ ev. Then if we suppose that p is of the same order as p', we find that for both MgO and BaO, χ is nearly zero. We cannot of course predict a definite value, but it is unlikely that χ is greater than 1 ev., and it could have a small negative value.

In the case of ZnO, it is fairly certain that χ has a positive value, and the probable limits are from $+1$ to $+3$ ev. It is doubtful whether the excitation level is as high as the zero line, but it is certainly higher than the bottom of the conduction band. Our conclusions differ from those of Seitz (1940), since he did not include either the polarisation terms or the width of the negative ion band. This applies equally to the case of ZnS, which Seitz considered in 1938 and 1940. Here χ is of the order $+5$ ev. In the case of ZnS, as in all the compounds dealt with in this section, the separation between the bands is considerably smaller than the first excitation energy, so that we expect photoconductivity in the tail of the first optical absorption band. This is observed in ZnS.

§ 10. THERMIONIC EMISSION

We have found that for all the oxides in §§8 and 9 except ZnO, the predicted value of χ is nearly zero, and that the maximum value which appears possible increases from 0·5 to 1·0 ev. in the order Ba Sr Ca Be Mg, though in each case the value might be negative. There is some experimental evidence to show that the value is 0·3 for BaO and SrO. Now, when a metal impurity atom is added with an electron energy-level at a depth 2ϵ below the bottom of the conduction band, the oxide becomes a semi-conductor with activation energy of conductivity ϵ. The thermionic work function is nearly $\chi + \epsilon$. We expect the value of ϵ to be related to the ionisation potential of the metal atom, and therefore to be larger, for example, for Be than for Ba. Thus if the value of χ were the same for BeO as for BaO, we should nevertheless expect the thermionic work function of BeO containing Be to be greater than that of BaO containing Ba, as is observed. If, however, Ba could be introduced into one of the other oxides whose χ was similar to that of BaO, it might be possible to obtain the same thermionic work function as that of BaO, and of course if one of the other oxides had a smaller χ than that of BaO, a correspondingly smaller work function might be obtained. Intermediate values would be expected with a mixture of oxides in the case where they form a solid solution. These considerations apply to the mixtures of BaO and SrO, and of BaO, SrO and CaO which are commonly employed as coatings for oxide cathodes. In such mixtures, solid solutions are formed, and when free metal is produced by a reduction or electrolytic process, the resultant metal is mainly Ba, since the heat of formation of BaO is less than that of the other oxides. It is known that the work function of such coatings is of the same order as that of BaO, and may be slightly less, although the work functions of SrO containing Sr and CaO containing Ca are considerably greater. This is understandable according to the above argument. We therefore deduce that the values of χ for SrO and CaO are similar to the value for BaO. An accurate comparison will be possible when the work functions of the mixed oxides have been measured accurately over a range of composition. Mixtures of BaO with MgO and BeO do not appear to have been studied.

§ 11. SECONDARY EMISSION

It is well known that all the oxides considered except ZnO have high secondary emission. Bruining and de Boer (1939) suggested that in the insulators with high secondary emission, the first empty band is situated above the zero level, so that when the incident primary electron causes a transition, the secondary raised from the full negative ion band can readily escape. The present results indicate that though the whole of the conduction band is probably not situated above the zero level, the first allowed transition always raises the electron above the zero level, which is in agreement with the general argument of Bruining and de Boer's theory. The selection rules for transitions caused by electron bombardment are similar for primary energies of several hundred volts to those for optical transitions, according to Wooldridge (1939). There does not appear to be any information concerning ZnO, but since χ is probably nearly 3 ev., and the excitation level is near zero, it is not likely to be as good a secondary emitter as the other oxides.

REFERENCES

DE BOER, 1935, *Electron Emission and Absorption Phenomena* (Cambridge), p. 241.
BRUINING and DE BOER, 1939, *Physica*, **6**, 834.
FESEFELDT and GYULAI, 1929, *Nachr. Wiss. Gottingen*, 229.
LOZIER, 1934, *Phys. Rev.*, **46**, 268.
MAYER and MALTBIE, 1932, *Z. Phys.*, **75**, 748.
MOTT, 1938, *Trans. Faraday Soc.*, **34**, 500.
MOTT and GURNEY, 1940, *Electronic Processes in Ionic Crystals* (Oxford).
MOTT and LITTLETON, 1938, *Trans. Faraday Soc.*, **34**, 485.
NISHIBORI and KAWAMURA, 1940, *Phys. Math. Soc. Japan*, **22**, 378.
O'BRYAN and SKINNER, 1934, *Phys. Rev.*, **45**, 370.
O'BRYAN and SKINNER, 1940, *Proc. Roy. Soc.*, A, **176**, 229.
SEITZ, 1938, *J. Chem. Phys.*, **6**, 454.
SEITZ, 1940, *Modern Theory of Solids* (McGraw-Hill), p. 448.
SHERMAN, 1932, *Chem. Rev.*, **11**, 93.
WOOLDRIDGE, 1939, *Phys. Rev.*, **56**, 562.

Energy Levels in Oxide Cathode Coatings

BY D. A. WRIGHT

(Communication from the Research Staff of the M.O. Valve Co. Ltd. at the G.E.C. Research Laboratories, Wembley, England)

MS. received 1 *March* 1947; *in revised form* 21 *May* 1947

ABSTRACT. The energy level diagram is considered in the case of a Ba/SrO emissive cathode coating, and it is concluded from the results the preceding paper that the process of activation consists in building up a concentration of free barium in the coating, and in providing barium at the interface between core and coating. The work function of the activated coating without adsorbed barium on its outer surface cannot be much greater than the observed value of 1 ev., and may be no greater, so that such an adsorbed layer, if present, has only a small effect on the work function. The possibility of increase in emission under the influence of ultra-violet light or electron bombardment is briefly discussed.

§1. THE METAL-COATING CONTACT AND THE WORK FUNCTION

THE oxides discussed in the preceding paper have not been investigated directly as regards energy levels, but indirect information is available from other studies. In the present paper we are concerned with BaO and SrO, which in the mixed oxide form are employed as thermionic emitters. The double oxide Ba/SrO is applied to a metallic base, and in the highly emitting state is a semi-conductor activated by the presence of excess barium. In studying the mechanism of emission, measurements have been made of the thermionic work function, the electrical conductivity, the photoelectric behaviour and the secondary emission. There has been considerable variation in results obtained, as a result of which the details of the mechanism remain uncertain.

Three types of picture have been proposed: (*a*) According to de Boer (1935) barium atoms are adsorbed on the outer surface of the coating, and can lose their electrons by thermal ionisation. The ionisation energy of the adsorbed atom is less than that of a free atom. The emitted electrons are replaced by a flow of electrons through the coating. (*b*) According to Reimann (1934) the electrons are emitted from the conduction band of the semi-conductor. It is supposed that an adsorbed layer of barium reduces the effective work function compared with that of the "clean" coating, precisely as when such a layer is adsorbed on the surface of a more electro-negative metal. (*c*) It is possible that the emission is the inherent emission of the coating as in (*b*), but that the work function of the "clean" coating is sufficiently low to lead to high emission. In this case it would be possible for high emission to be obtained in the absence of an adsorbed layer of barium. In all three theories it is necessary to suppose that excess barium is freed in the coating, thus raising its conductivity compared with that of the pure oxide. This barium is freed by electrolysis or by chemical reaction with reducing agents in the core metal or coating. At the interface between core and coating a potential barrier is to be expected, whose height we shall assume for the moment to be equal to the difference between the work functions of core metal and coating. The barrier may be of the Schottky type, or a layer may be present between core and coating whose resistance is higher than that of the coating. These barriers have been discussed by the writer (1947), and evidence for their presence has been given by Fineman and Eisenstein (1946).

The thermionic work function of the coating is the height OE in figure 1, where O is the potential energy at infinity, A is the bottom of the conduction band in the crystal, and C the top of the full band as in the preceding paper. D is now an added level due to the Ba impurity, and the level E is approximately half-way between A and D.

When the coating is in equilibrium with the core metal, the levels in the metal are raised until the top of the Fermi level in the metal F is at the same height as E, figure 2. Thus PF is the difference between the two work functions. Now experiments by the writer have indicated that the value of PF is 0·7 ev. when the core metal is nickel containing a small proportion of magnesium. The work function of Ni is 5·0, thus if the magnesium is unimportant in influencing the work function, the coating work function must be 4·3. Now there is good experimental evidence that the thermionic work function of the activated coating is of the order 1 ev. This appears at first to rule out picture (*c*) above, and to require either (*a*)

in which the energy levels are not directly involved, or (*b*) in which an adsorbed layer of barium on the outer surface reduces the effective work function. The energy level system is then as in figure 3.

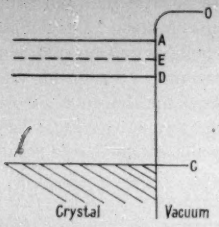


Figure 1. Energy levels in excess semi-conductor.
C is the top of the upper fully occupied band.
A is the bottom of the conduction band.
D is the added impurity level.
E is half way between A and D.

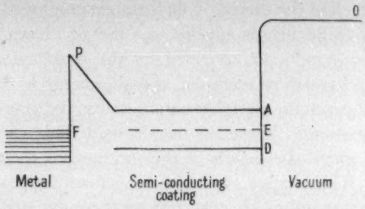


Figure 2. Excess semi-conductor in contact with metal. Case where semi-conductor surface is "clean".
A, E and D are levels as in figure 1.
F is the top of the Fermi levels in the metal.
PF is the difference in work functions of metal and semi-conductor.

This type of diagram was discussed by Reimann (1934), except that he considered the barrier at the metal-coating interface to be thin enough for tunnel effect penetration. The writer has indicated elsewhere (Wright, 1947) that in many cases the barrier is too thick for this to occur. Here the energy level diagram discussed was as in figure 3, i. e. it was supposed that the adsorbed barium layer might be present at the outer surface.

There were, however, clear indications that the behaviour of the coating was determined more by the conditions at the metal-coating interface than by the conditions at the outer coating surface. Thus the time-decay effects in thermionic emission current and in conduction current were related and were both associated with the core-metal boundary. Thus if mechanism (*a*) is applicable, we suppose that changes at the interface affect the conductivity, and through it the emission, because of the electron supplementation process. It seems necessary to attribute the decay effects to electrolytic flow, which led the writer (1947) to the tentative proposal that positive ions are adsorbed at the metal-coating interface. Alterations in their distribution then leads to the decay effects in emission and conduction current. Now it is important to note that this possibility alters our conclusion above that mechanism (*c*) is ruled out. Thus if we assume that the work function of the "clean" oxide without adsorbed barium has the observed value of 1·0, then with PF = 0·7 we require the work function of the nickel with an adsorbed layer

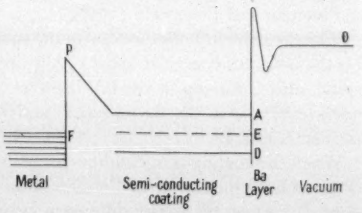


Figure 3. Excess semi-conductor in contact with metal. Case where adsorbed electro-positive layer is present on semi-conductor surface. Levels as in figure 2.

of barium to be 1·7, which is quite a possible value. Thus we have one extreme possibility consistent with (*b*), that the work function of clean Ba/SrO is 4·3, but an adsorbed layer of Ba on the outer surface reduces the effective value to 1·0, or the opposite extreme consistent with (*c*), that the work function of the coating is 1·0, but an adsorbed layer of Ba at the metal-coating interface reduces the height of PF from 4·3 to 0·7. The decay phenomena indicate that (*b*) is not a true picture. The alternatives are (*c*), a state intermediate between (*b*) and (*c*), or de Boer's mechanism (*a*).

It should be noted that the experimental evidence concerning the work function does not in most cases throw any light on this question. The fact is established that the thermionic work function of BaO or of Ba/SrO is nearly 1 ev., but most results give no indication whether the true situation resembles figure 2 or figure 3. In the case of the work of Nishibori and Kawamura (1940), referred to in the preceding paper, it was found that the value of χ remained at 0·3 ev. during activation, although the thermionic work function decreased by 0·4 ev. This indicated that the only effect of the freeing of Ba during the activation lay in increasing the concentration of impurity levels D and consequently raising the level of E. Thus this evidence was strongly in favour of figure 2. It is, however, possible to argue that surface Ba was present at the beginning of the activation in these experiments, and that its concentration did not vary during activation. If this were the case, the evidence would not be conclusively in favour of figure 2. It seems unlikely in fact that a surface concentration could be maintained constant, independent of the internal concentration; however, in view of this doubt, the calculations in the preceding paper were carried out. These show that the value of χ is not greater than 0·6 for BaO without a surface layer of Ba. Since the experimental values for the activation energy of conductivity AE lie between 0·5 and 1·0 ev., we conclude that the thermionic work function in the absence of adsorbed Ba cannot be greater than 1·6. This rules out the extreme case referred to above where the work function of clean oxide is 4·3 ev., and the adsorbed layer of Ba is very important in reducing the work function after activation to 1·0 ev. It permits the other extreme case corresponding with figure 2, and theory (*c*), in which the work function of the clean oxide can become 1·0 after activation, without the presence of adsorbed Ba. It also permits intermediate cases in which, with an adsorbed layer of Ba present, the work function is slightly lower than in the absence of such a layer, i.e. lower by a few tenths of an electron volt. This would correspond with a form of figure 3 in which the depression of the work function due to the adsorbed layer is small.

We conclude therefore that the thermionic emission from BaO or Ba/SrO is emission from the conduction band of the semi-conductor, the bottom of which is situated not more than 0·6 ev., and probably only 0·3 ev. below the zero level. The process of activation consists in building up a sufficient concentration of free Ba in the oxide to lower the work function to 1 ev. by raising the level E, and in providing a layer of adsorbed Ba at the coating-core interface. At the outer coating surface, while Ba must be present at least intermittently, since it evaporates from the oxide, it does not play a very important part in lowering the work function to the observed value of 1 ev. in the activated cathode. There is an indication that it may play no part at all, i. e. that the clean oxide with an adequate internal

concentration of free Ba has a work function of 1 ev. If the adsorbed layer of Ba does play a part, it can only be small, since the work function of the clean oxide cannot be much more than 1 ev. We may note here that the conclusions of Huber and Wagener (1942) were that there is no electron diffraction evidence to support the view that a layer of adsorbed Ba is present on an activated cathode.

§ 2. INFLUENCE OF LIGHT AND ELECTRON BOMBARDMENT

When electrons are raised into the conduction band of a semi-conductor, for example, under the influence of light of sufficiently short wavelength, the conductivity is increased. The increase in the number of electrons in the conduction band which causes this should also increase the thermionic emission from the semi-conductor. Apart from this short-time effect, there may be a photochemical decomposition, as in the alkali and silver halides, where loss of an electron from a negative ion can lead to the loss of a negative ion from the crystal. If there is a corresponding effect in the oxides, there should be an increase in conductivity and in thermionic emission from the oxide under the influence of the light, if the oxide is not initially fully activated.

There is some evidence that one or both of these processes may occur in oxide coatings (Case, 1921; Merritt, 1921; Crew, 1926; Newbury, 1929; Bodemann, 1929; Ramanadoff, 1931). In these papers it was found that there was a rise in photoelectric response in the near ultra-violet with temperature. It was at first thought that there was an increase in photo-electric emission with temperature; it later appeared, however, that the true effect was an enhancement of the thermionic emission under the influence of light of sufficiently short wavelength. There was no effect above 3000 A., a small effect from 3000 to 2500 A., and a considerable increase near 2500 A. By analogy with the results on the photo-conductivity of silver halides, it can be deduced from these observations that the range 5 to 4 ev. lies in the tail of the first optical absorption band, whose peak will therefore occur at considerably higher energy. In the case of electron bombardment of an oxide, similar effects might be expected. Johnson (1946) obtained effects which he interpreted as the enhancement of emission from poorly activated oxide cathodes under electron bombardment, though the interpretation is doubtful in view of the later results of Pomerantz (1946), who found a true increase in secondary emission with temperature. As regards the photo-chemical effect, both Headrick and Lederer (1936) and Jacobs (1946) have obtained evidence for the liberation of oxygen from oxides under electron bombardment. If this does occur in oxide cathodes, it is possible that continued bombardment of an activated cathode would lead to the formation of an excess of free metal and hence of metal aggregates, though such an effect would be temperature dependent. Their formation at low temperature might explain the observed decay in secondary emission with time, since metal aggregates may decrease the secondary emission both by replacing the oxide and by providing electron traps. At higher temperatures, for example 700 or 800° C., the excess metal would evaporate, and the loss of oxygen and metal would lead to disappearance of coating at a greater rate than in the absence of electron bombardment. Such an effect is observed in cavity magnetrons where some of the emitted electrons return to bombard the cathode.

REFERENCES

BODEMANN, 1929, *Ann. Phys., Lpz.*, **3**, 614.
CASE, 1921, *Phys. Rev.*, **17**, 388.
CREW, 1926, *Phys. Rev.*, **28**, 1265.
DE BOER, 1935, *Electron Emission and Adsorption Phenomena* (Cambridge), p. 362.
FINEMAN and EISENSTEIN, 1946, *J. Appl. Phys.*, **17**, 643.
HEADRICK and LEDERER, 1936, *Phys. Rev.*, **50**, 1094.
HUBER and WAGENER, 1942, *Z. tech. Phys.*, **23**, 1.
JACOBS, 1946, *J. Appl. Phys.*, **17**, 596.
JOHNSON, J. B., 1946, *Phys. Rev.*, **66**, 352.
MERRITT, 1921, *Phys. Rev.*, **17**, 525.
NEWBURY, 1929, *Phys. Rev.*, **34**, 1418.
NISHIBORI and KAWAMURA, 1940, *Phys.-Math. Soc. Japan*, **22**, 378.
POMERANTZ, 1946, *Phys. Rev.*, **70**, 33.
RAMANADOFF, 1931, *Phys. Rev.*, **37**, 884.
REIMANN, 1934 *Thermionic Emission* (Chapman and Hall), p. 228.
WRIGHT, D. A., 1947, *Proc. Roy. Soc.*, A, **190**, 394.

The Experimental Basis of Electromagnetism: Part II—Electrostatics

BY N. R. CAMPBELL AND L. HARTSHORN

MS. received 28 *March* 1947

ABSTRACT. This paper continues the argument of a previous paper with the same title. It explores the foundations of the science of electrostatics as practised in the modern laboratory, admitting as evidence only those experiments that experience has proved to be practicable. The basis of the subject is found in alternating currents and the laws of capacitance, which lead to the conception of electric potential energy. The various experiments on the mechanical forces between electrified bodies fit in well with this conception, and we are led to the conclusion that the soundest procedure in the investigation of any system of conductors and dielectrics is to represent it by its equivalent capacitance network. Coulomb's Law and point charges play no part in our scheme, but they are discussed because of the prominent part that they play in classical theory. The familiar equations for point charges, which are the premises from which that theory starts, can be regarded as definitions of what is meant by " point charge ", or as approximations to the laws governing the forces on small charged bodies, which become true in a limiting case. One incidental advantage of our treatment is that the somewhat elusive notion of " earth " in the usual expositions is replaced by more definite conceptions.

§1. ALTERNATING CURRENTS

THE general principles to be followed in this enquiry were outlined in a previous paper (1946), here called Part I, which established the magnitudes characteristic of D.C. circuits, viz. current, voltage, resistance etc. Our present object is to show that the working principles of electrostatics as practised in the modern laboratory can be soundly based on real experiments, and in this way to indicate the true nature of the experimental foundations of the subject. It is a commonplace among experimental workers that the magnitudes that appear in electrostatics are today almost invariably measured by alternating current methods,

which are far more accurate than those used by the pioneers in the subject; we naturally therefore look to alternating currents for our basic conceptions.

In one of the circuits from which we started, let the battery be replaced by an A.C. generator (say audio-frequency). Current indicators in series will not generally read the same current, according to their D.C. calibration; nor will voltage indicators in parallel read the same voltage. If the indicators are of the kind generally employed in modern practice—but not necessarily otherwise—most of the direction-sensitive indicators will read nothing at all, while the reversible indicators will continue to agree among themselves. If we had started from such an arrangement, we should simply have ignored the direction-sensitive indicators as not indicators at all.

But, even then, if we had tried to calibrate the reversible indicators by addition in the manner described in Part I, we should have failed, because the necessary laws of addition are not true in general. They are true if the circuit elements are limited to a narrow class (that of "pure" elements all of the same kind, see §2); and this fact is used in the calibration of A.C. instruments according to the best modern practice. But, though we might by such methods measure independently magnitudes corresponding to current and voltage, it would be doubtful whether we should be wise to call them by those names. For there would be no negative currents or voltages; and the laws from which "current" and "voltage" derive so much of their significance would not be true; neither the sum of the currents flowing to a node nor the sum of the voltages round a circuit would be zero.

However, if we include cathode-ray oscillographs among the indicators, we shall find that they, though direction-sensitive, continue to give an indication; but they indicate that the state of the system is not constant, but is varying cyclically. If we calibrate these instruments by D.C. for current and voltage, and by some form of stroboscope for time, and if we then regard only those values of current and voltage that are simultaneous, we shall find that simultaneous current and voltage possess over a wide range the properties of direct current and voltage; in particular they obey the laws from which these terms derive their significance. Moreover, we shall find that the behaviour of all the instruments that give steady indications in an A.C. circuit can be completely explained in terms of their reactions to different direct currents and of the presence of some kind of inertia that causes them to read a mean value of a rapidly varying influence. Electrostatic and dynamometer instruments agree among themselves because they all read nearly the same mean; rectifier instruments diverge because they read a different mean.

These are the considerations that justify the conceptions of instantaneous and RMS current, voltage and power, and that underlie the calibration of RMS instruments in terms of D.C. in standardizing laboratories. Certain elaborate precautions have to be observed; but since they do not involve any of the laws we are about to discuss, they need not detain us.

§2. PURE RESISTORS AND PURE CAPACITORS

When the A.C. generator is substituted for the battery, none of the laws concerning circuit elements set forth in Part I remains generally true. In particular, if equality of impedance or admittance means equivalence in respect of a reversible indicator in an A.C. circuit, impedance is not generally additive in series

or admittance in parallel; consequently impedance and admittance are not, like resistance and conductance, measurable by addition. But it is possible by certain tests to select, from the whole class of elements that may form part of an A.C. circuit, two sub-classes that possess characteristics that are additive and are therefore measurable independently; they will be termed respectively "pure resistors" and "pure capacitors". (In a later part of this enquiry it will be pointed out that there are no even approximately pure inductors.)

This statement needs some explanation. There is no circuit element that satisfies completely all (or perhaps any of) the tests that will be prescribed in their most stringent form. But, if elements are selected that very nearly satisfy all the tests, certain simple laws are found to be very nearly true of them; departures from the laws are concealed by experimental error in all but the most accurate experiments, and even in those are distinguishable from experimental error only by careful analysis. Accordingly we define a pure element as one of which these simple laws are true, and attribute departures from purity in actual elements to a combination of elements of different kinds, each of which would be pure in isolation. We find that the behaviour of actual elements can be explained on this hypothesis. The statements that follow must be interpreted in view of this procedure which, of course, is adopted with minor modifications in many branches of physics.

The distinction between pure resistors and pure capacitors is that the former do, and the latter do not, pass current in a D.C. circuit. (Pure capacitors thus violate the most fundamental property of D.C. circuits, namely that the path of the current lies wholly in conductors.) One test of purity applicable to both sub-classes is that the laws of addition in parallel and series must actually be obeyed. This test, since it involves a plurality of elements, can be applied only to the sub-class as a whole, not to individual members of it. It might fail completely for this reason; for impure elements might be additive, if they were all impure in the same way. Actually the test is very useful, because impure elements usually differ in their impurity. Further, the test, applied to capacitors, limits the sub-class to what we shall call (for a reason that will appear presently) "closed" capacitors (in common parlance, screened condensers). Other elements that are called pure capacitors according to convention—which we shall follow—are not additive.

Accordingly it is desirable to have other tests. Another class test is that the equivalence of two elements of the same sub-class should be independent of the indicator (i.e. of the nature of the mean it indicates) and of the generator (i.e. of its frequency and wave form). But this test, though satisfied by pure capacitors, is not fully satisfied by pure resistors; in ordinary language, their resistance varies slightly with the frequency.

There is a very perfect individual test for pure resistors. It is that the current through the resistor and the voltage across it, measured by cathode-ray oscillographs, must be "in phase"; in particular, the maxima and minima of current must be respectively simultaneous with the maxima and minima of voltage; this test can be applied without measuring anything, and involves only the relations of greater and less. The corresponding test for pure capacitors is less easily applied and its consideration therefore is postponed. The best individual test for capacitors is that which distinguishes them from resistors, namely that they pass no

direct current; but this is not conclusive, since it does not include dielectric loss and series resistance or inductance. The class tests for capacitors are therefore more important than the class tests for resistors.

The magnitudes characteristic of pure resistors that are additive in series and parallel turn out to be the same as D.C. resistance and conductance, at least at low frequencies; no separate name for them is required. On the other hand, new names are required for the corresponding magnitudes characteristic of capacitors; we shall call them elastance and capacitance. In modern terminology these magnitudes are distinguished from impedance and admittance; but, from our present standpoint, when we are considering independent measurement alone, they are the same magnitude according to our meaning of that expression (Part I, p. 645). For, whatever the frequency and wave-form, the ratio of the impedances or admittances of two pure capacitors is the same as the ratio of their elastances or capacitances.

The measurement of capacitance by the parallel connection of pure closed capacitors plays as important a part in the most refined calibration of pure capacitors as does series connection in the calibration of D.C. resistors. The measurement of capacitance is slightly less accurate than that of resistance, because (for a reason that will appear presently) the conditions in which the law of addition is true cannot be realized so precisely. For the same reason, and perhaps for others also, elastance plays a much less important part than capacitance in actual measurement; but it should be noted that, in so far as it is an independently measurable magnitude, it must be the reciprocal of capacitance, if the same standard is used to define both units.

§ 3. CAPACITANCE AS A DERIVED MAGNITUDE

If we measure by cathode-ray oscillographs both the current I flowing through a pure capacitor and the voltage V across it, then we find that I is proportional to the simultaneous dV/dt, and that this relation is independent of the particular generator used and of the way in which V varies. The constant of proportionality depends on the capacitor, and is therefore a derived magnitude; we may write the law

$$I = S \,.\, C \,.\, dV/dt \qquad \ldots\ldots (3.1)$$

C turns out to be the same magnitude as the capacitance measured by addition, when the capacitor is closed and obeys the law of addition. It we choose suitably, with respect to the units of current, voltage and time, the closed capacitor to which unit capacitance is to be assigned, the scale factor S may be made 1. This is the convention always adopted; we shall assume its adoption in what follows.

Equation (3.1) is true even if the capacitor is not closed, so long as it satisfies the tests of purity other than that of addition. Accordingly, as indicated already, we shall regard capacitors obeying (3.1) as pure, even if they are not closed. It follows from (3.1) that, if there is a definite phase difference between I and V—which implies that both of them are sinusoidal—then it is 90°. This corresponds to the phase test for pure resistors; and since a phase difference of 90° implies that the maxima of one quantity coincides with the minima of the other, can be applied without measurement. But if the capacitor is not entirely pure, and the phase difference not exactly 90°, the distinction between a leading and a lagging current,

which will turn out to be important, will arise. It is therefore doubtful whether the test can actually be applied in a form in which it differs from the test whether (3.1) is true.

(3.1) cannot be established with great accuracy by oscillograph measurements, if only because the deduction of a derivative from a set of disparate values is always liable to considerable error. But on the assumption that (3.1) is true, it is possible to devise bridge networks fed by generators of sinusoidal wave-form (so that the relation between V and dV/dt is known) that permit C to be determined, both relative to other C's and to the conductances of pure resistors. The consistency of the results obtained with such networks confirms with high accuracy the truth of (3.1), and the identity of C in that law with capacitance measured by addition in closed capacitors.

Just as the most refined measurement of resistance depends both on the law of addition and on Ohm's law, so the most refined measurement of capacitance depends both on addition and on (3.1). Bridges are also useful in judging equality in independent measurement; it is worth noting that, when they are used for this purpose, the wave-form of the generator need not be sinusoidal.

§ 4. CAPACITANCE AND GEOMETRICAL FORM: PERMITTIVITY

On examination of their structure, capacitors prove to consist essentially of a pair of conductors ("plates") separated by a gap filled with a non-conducting material, or dielectric and generally narrow compared with the dimensions of the plates. When the plates approximate sufficiently nearly to concentric spheres, coaxial infinite cylinders or infinite parallel planes, experiment shows that there are simple relations between capacitance and geometrical form. In the usual notation they are

$$C = K_1 \,.\, r_1 r_2/(r_2 - r_1) \qquad \text{(spheres)} \qquad \ldots\ldots (4.1)$$

$$C = K_2 l/(\log r_2 - \log r_1) \qquad \text{(cylinders)} \qquad \ldots\ldots (4.2)$$

$$C = K_3 A/d \qquad \text{(planes)} \qquad \ldots\ldots (4.3)$$

K_1, K_2, K_3 vary with the dielectric; each is therefore a derived magnitude requiring a scale factor. But it turns out that they all vary together, so that—within experimental error—$K_1 = 2K_2 = 4\pi K_3$. Accordingly, if we write

$$K_1 = S \,.\, \kappa; \quad K_2 = \tfrac{1}{2} S \,.\, \kappa; \quad K_3 = 1/4\pi \,.\, S \,.\, \kappa. \qquad \ldots\ldots (4.4)$$

a single magnitude κ (permittivity) characteristic of the dielectric and a single scale factor will suffice. Experiment shows further that the independence of permittivity and geometrical form is complete, so that if a value for κ is assigned by convention to one dielectric, it can be determined for any other by measuring the capacitance of a capacitor of any form, filled first uniformly with the standard dielectric and then uniformly with the other.

Somewhat more complicated formulae have also been established experimentally when the dielectric, though not uniform, is distributed in uniform layers. They involve different κ's for the different layers, but they need no special consideration here.

All these formulae are special cases of a general rule for determining capacitance from the geometry of the capacitor and the nature of the dielectric. The form in

which it is given here depends historically on laws not mentioned so far. But the discovery of the rule by guesses based on empirical laws such as (4.1), (4.2), (4.3) would have been no more remarkable than the discovery of the Ampère-Neumann rule (Part I, (11), (11.1)) by a similar process.

The rule depends on a geometrical theorem that can be stated with sufficient accuracy and generality for our purpose thus*:—

Let, $s_1 \ldots s_k, \ldots s_n$ be a set of closed non-intersecting surfaces. Let each point outside these surfaces be associated with a quantity κ, characteristic of it. Then it is possible to find a variable v having the following properties:—

(1) v is constant over each surface s;

(2) at all points outside the surfaces s, v is continuous and finite or zero, being constant at all points sufficiently distant from the surfaces;

(3) at each point at which κ is continuous, v satisfies the differential equation

$$\frac{d}{dx}\left(\kappa\frac{dv}{dx}\right)+\frac{d}{dy}\left(\kappa\frac{dv}{dy}\right)+\frac{d}{dz}\left(\kappa\frac{dv}{dz}\right)=0; \qquad \ldots\ldots(4.5)$$

(4) at any surface at which κ changes discontinuously, $\kappa \,.\, dv/d\boldsymbol{n}$ has the same value on either side of the surface, $\boldsymbol{n}$ being the normal to it.

It follows from the properties of v that, if determinate values v_1, v_2 are assigned to two of the surfaces s_1, s_2, while the values for the other surfaces are left undetermined, the quantity

$$\int \kappa \,.\, dv/d\boldsymbol{n} \,.\, dS = 4\pi\phi_e, \qquad \ldots\ldots(4.6)$$

where dS is an element of surface, $\boldsymbol{n}$ is the outward normal, and the integral is taken over the surface, has the same modulus but opposite signs for the surfaces s_1, s_2, and is zero for any other of the surfaces.

In applying this purely geometrical theorem to a capacitor, s_1, s_2 are identified with the surfaces of the plates, each plate including all the conductors connected conductively to it. v_1, v_2 are chosen arbitrarily. The value of κ associated with each point is the permittivity of the medium occupying it. If any point is occupied by a conductor insulated from both plates, κ for it is infinite; v is then constant over the surface of that conductor and ϕ_e for it is zero. Then v is in general determined everywhere; the quantity ϕ_e is everywhere proportional to $(v_1 - v_2)$; and the rule for capacitance is

$$C = S\left|\frac{\phi_e}{v_1 - v_2}\right| \qquad \ldots\ldots(4.7)$$

where ϕ_e refers to the plates and S is the scale factor of (4.4).

This is the most general relation between capacitance and geometrical form; formulae (4.1), (4.2), (4.3) are special cases, but they do not provide a satisfactory test of it, because the capacitors to which they refer cannot be realized fully. However, they can be realized approximately; the rule is then found to give a "correction" that agrees with measurement. If the rule is applied to capacitors

* The reader will note that it is assumed that there are no surface or volume charges in the dielectric. The modification required if there are such charges is considered below in § 16. But it should be pointed out that the same assumption is made in § 5. He may note further that no reference is made to the vector D, of which ϕ_e is the flux. The reason for this omission will appear in later parts of this enquiry.

differing widely from these simple forms, mathematical difficulties often prevent the prediction of a definite value with which measurement can be compared. The direct evidence for the accurate and universal validity of the rule is therefore not complete; but it is very strong—stronger than that for any other "electrostatic" relation.

§ 5. THE FIELD EQUATIONS: ELECTRIC FIELD STRENGTH

In § 4 v is a mere mathematical variable, convenient (but perhaps not absolutely necessary) in arriving at formulae (such as (4.1–3)) applicable to experiment; it does not appear at all in the formulae. But further inquiry shows that v, or more accurately grad v, has great physical significance.

In Part I the magnitude H was introduced to express the fact that there is a group of phenomena, including the deflection of a compass needle or of a cathode ray, that are characteristic of the space about a current-carrying circuit, and are determined by the product of the current in that circuit by a geometrical function G. There is another group of phenomena similarly characteristic of the space about conductors between which voltages are maintained; like the "magnetic phenomena", they form a group, because if one of them is the same at two points in such a space (or spaces), then each of them is the same at those points. The group includes again the deflection of a cathode-ray; it includes also the Kerr electro-optical effect and the breakdown of a dielectric.

Suppose that we identify the surfaces of the conductors with the surfaces s, assign values of v to them so that $v_k - v_l$ is the voltage V_{kl} maintained between s_k and s_l, and calculate v throughout the space between them by means of the theorem of § 4. Then we find that the phenomena are determined by grad v in the sense that they are the same wherever grad v is the same. Accordingly

$$\mathscr{E} = \text{grad}\, v \qquad \ldots\ldots (5.1)$$

is a defined magnitude determining these electric phenomena in the same way as H determines the magnetic phenomena; it may be fitly called the electric field strength. If there are only two conductors between which a voltage is maintained, if V is that voltage, and if the dielectric is uniform, the similarity between E and H may be made even more evident by writing

$$\mathscr{E} = VF, \qquad \ldots\ldots (5.2)$$

where F, like G, is a function depending only on the geometry of the system.

In view of these facts we shall term the propositions of § 4 collectively the *field equations*. Further, we shall assume—for this is obviously the easiest way—that the v's are to be assigned to the conductors in accordance with the voltages, by assigning $v = 0$ to one conductor (zero plate) and to the others their voltages to this conductor.

In some circumstances—it is unnecessary here to inquire what they are—it is possible to establish that the voltage between the zero plate and a probe introduced into the space between the conductors is equal to v, calculated from the field equations. Then v, as well as grad v, has physical significance. Measurements of probe voltages are inaccurate and have little evidential value. On the other hand, $\mathscr{E}$ can be determined with considerable accuracy from the deflection of a cathode-ray, by the use of a theorem that will be examined in § 15. It is possible that the

future development of electron optics may provide evidence for the field equations more stringent than that derived from the measurement of capacitance; but for the present they must be based primarily on such measurement.

v of this section is closely related to the v that was introduced in Part I to explain the properties of conducting media. It follows from the field equations and from the properties of v set forth in § 13 of Part I that, if the dielectric of a capacitor is replaced by a medium whose conductivity is everywhere proportional to the permittivity of the dielectric, and if the plates are still maintained at the same voltage, then the v of the conductor is everywhere the same as the v of the capacitor. This relation is sometimes used to determine the distribution of v in a capacitor of complicated form. But it is more important because there is no sharp distinction between dielectrics and conductors; there are media in which it would be possible to determine both equipotentials as indicated in § 13 of Part I, and the electric field strength by one of the group of phenomena mentioned earlier in this section; more generally we find that the properties of most dielectrics can be explained by attributing to them both permittivity and conductivity. It is an experimental fact of great practical importance that, in isotropic media, the distribution of v is the same whether it is derived from conductive or from electrostatic phenomena. Perhaps the strongest evidence on this point is the well-known fact, frequently established with great precision, that the "loss tangent", and therefore σ/κ, for any uniform medium is independent of geometrical form.

§ 6. CLOSED CAPACITORS: SCREENS

The exceptions to the statement in § 4, that v is determined everywhere by the assignment of v_1, v_2, are very important.

(*a*) If a conductor, insulated from both plates, encloses one but not the other, then v is completely indeterminate unless something is added to the rule. The obvious addition, based on the fact that ϕ_e for a conductor insulated from both plates is zero, is that, if ϕ_e' is the value of ϕ_e for the enclosed plate, then $-\phi_e'$ is its value for the inner surface of the enclosing conductor, $+\phi_e'$ for its outer surface, and $-\phi_e'$ for the unenclosed plate. The experimental facts that justify this addition will be stated later. With this addition, which will be assumed in what follows, v in this case is determinate everywhere.

(*b*) If one of the plates completely encloses the other, then the assignment of v to the enclosed plate and to the inner surface of the enclosing plate, though it determines v everywhere within the enclosing plate, does not determine v anywhere outside it. The corresponding experimental fact is that changes in the voltages between conductors inside the enclosing plate (including the inner surface of that plate) produce no change of voltage between any pair of conductors outside the enclosing plate (including the outer surface of that plate), and no change of the field strength outside that plate; and conversely, interchanging "inside" and "inner" with "outside" and "outer".

This deduction from the field equations can be tested with great sensitivity; since it is true, some variations from the rule of capacitance given by the field equations as stated in § 4, which might otherwise be possible, are excluded. But the fact is even more important practically than theoretically. Thus, it is used in radio-frequency experiments when the observer, in order to isolate himself from

irrelevant disturbances, encloses himself and all his apparatus inside a metallic cage. (In view of later discussions, it is well to point out here that the efficiency of the cage for this purpose does *not* depend on its being connected to any other conductor, whether inside or outside the cage.)

Again, it makes it possible to limit the conductors that have to be taken into account in any electrostatic problem to a definite group. If one of a group of conductors encloses all the rest, then any conductor outside the group—and outside the enclosing conductor—can be ignored. This limitation and isolation by a "screen" is necessary both in theoretical arguments and in all experiments except the very roughest. Accordingly it will always be assumed hereafter, unless the contrary is stated explicitly, that the system under discussion is a group within and including a surrounding screen. Of course an isolating screen need not in practice be absolutely complete; the metallic cage above mentioned is often made of wire gauze or netting. The effect of a small aperture tends to zero with the ratio of its area to the whole area of the screen, and is often quite inappreciable.

Other closely related deductions from (*b*) are that the capacitance of a capacitor of which one plate (screen) completely surrounds the other is independent of any body outside it, and that the capacitance of a pair of such capacitors, with their screens and their plates respectively connected together, is the sum of their individual capacitances. If one plate does not surround the other, capacitance will not in general be additive in parallel, because the field between the plates of one capacitor affects the field between the plates of the other. This is, of course, why capacitors that obey the laws of addition were called "closed" in § 2.

If a capacitor could not be "closed", and obey the laws of addition, unless the screen *completely* surrounded the plate, a closed capacitor would be an unattainable ideal. For, in order that the plate of a screen-and-plate capacitor should be connected to anything outside the screen, e.g. another plate, there must be an aperture in its screen through which a lead may pass. The aperture itself, if small enough, may be unimportant; but the lead passing through it is another matter. It is a question of fact, not to be decided by anything stated so far, whether the departure from ideal closure, represented by the passage of a lead through the screen, produces a failure of the law of addition.

The answer depends, of course, on the sensitivity of the criterion of equality and on the size and location of the lead. If the most sensitive means for judging equality is used, and the capacitors have ordinary terminals, then the law of addition is found to be not *generally* true, although it is true for capacitors exceeding (say) 0·5 μF. in value, provided that reasonable precautions are taken in choosing and arranging the leads. It is, however, possible so to design capacitors that the leads connecting them together make no detectable contribution to their capacitance. For example, the plate terminals may take the form of sockets mounted in small apertures in the screen, so that the face of the socket is flush with the surface of the screen. If the sockets of two such capacitors register with one another when the screens are placed in contact, then they can be connected together by a double-ended plug, which is inserted in one socket before the capacitors are brought together, and which becomes totally enclosed when the connection is complete. When capacitors of this type are used, the law of addition can be verified with the highest accuracy over a very large range of values; indeed, it is probably the most

accurately known of all the quantitative laws of electrostatics. Moreover (see §12 below) there are methods of assigning capacitances to ordinary standard capacitors that are also additive in respect of the most sensitive criteria of equality; and these capacitances agree within experimental error with those calculated from the rule of capacitance. Accordingly there are actual capacitors completely equivalent to ideal closed capacitors, having no leads through their screens; and in basing the laws of electrostatics on the properties of ideal closed capacitances, we shall not be departing from our principle of relying only on real experiments and real conceptions.

There is a corresponding deduction from (*a*), namely that elastance is additive in series if, and in general only if, the plate common to the two capacitors completely encloses the other plate of one of the capacitors, and is completely enclosed by the other plate of the other. This condition is not easily realized, even if the leads are ignored, and elastance, even if it proved accurately additive in the presence of leads, would not be mportant. Actually the departures from accurate additivity of elastance are of the same nature as those just discussed; they are often concealed by experimental error; and even when they are not concealed, they may be abolished by the methods of §12 below.

§7. ELECTROMETERS

We now return to the main theme, namely the coordination of the laws of current circuits.

We note that electrostatic voltmeters and electrometers are essentially closed capacitors in which the plate is movable relative to the screen, so that their capacitance varies with the motion of the plate, and the voltage V to be measured is applied between plate and screen. (If such voltmeters were used to investigate (3.1), account would have to be taken of the variation of their capacitance, which is part of C, with V.) Experiment shows that their behaviour is consistent with the law

$$F_q = S \,.\, dC/dq \,.\, V^2, \qquad \ldots\ldots (7.1)$$

where F_q is the force tending to increase the coordinate q, the voltage being maintained independent of q by means of a battery or generator, and S (which is positive) is the same for all voltmeters that obey (7.1), so that it is a scale factor that does not need to be supplemented by a derived magnitude.

The only voltmeters in which the law can be established accurately are those in which C can be calculated accurately from the field equations; but, in some of these, it can be established as accurately as any of the laws previously mentioned in this Part. By assigning values to S in (4.4), (4.7), (7.1), and to κ of a standard dielectric, the units of C, V, I, κ can be fixed in terms of the units of length, area, time and force. They could actually be fixed by this method with an accuracy little less than that with which they are now fixed by other methods. In "electrostatic absolute units", κ is made unity for vacuum, S of (4.4), (4.7) are made unity, and S of (7.1) is made $\frac{1}{2}$. (The reason for this last choice will appear presently.)

§8. THE CHARGE IN A CAPACITOR

We have said that (7.1) is true only if V is made independent of q by connecting plate and screens through a battery or generator. Electrometers are sometimes

used idiostatically with plate and screen insulated from each other; then V varies with their relative position. We shall now show that the law corresponding to (7.1) for this case can be inferred from the laws already stated.

Suppose that a direction-sensitive ammeter is permanently connected in series with one of the terminals of the capacitor. While the capacitor is a circuit element, let us record I, the current flowing through the ammeter, as a function of time; and let us define a quantity Q, the *charge in the capacitor*, by

$$Q=\int_0^V I\,.\,dt, \qquad \ldots\ldots(8.1)$$

where the limits of the integral indicate that it is to be taken between an instant when the voltage across the capacitor is zero and an instant when it is V. Then from (3.1), with $S=1$

$$Q=CV. \qquad \ldots\ldots(8.2)$$

Comparison of (4.7) and (8.2) shows that

$$|Q|=|\phi_e|, \qquad \ldots\ldots(8.3)$$

accordingly in so far as Q, defined by (8.2), receives physical significance from the considerations of this section, ϕ_e receives the same significance.

When the terminals of a pure capacitor are not connected in a circuit, $I=0$. If (3.1) is true of these circumstances (an assumption not used so far), then V, and therefore by definition Q, must remain constant. Experiment shows that this condition may be approached very nearly in suitable circumstances; any change in V can be attributed to the presence of a conductor between the terminals whose conductance (representing a slight impurity) is so small that it cannot be detected in the range of frequencies within which C is most conveniently measured.

Let us now connect the terminals of the charged capacitor by a resistor of resistance R. Then, if Ohm's law is true when the appearance of a voltage across the resistor is not associated with the presence of a generator in the circuit of which it forms part (again a new assumption), we must have

$$I=-V/R; \qquad dI/dt=(-1/R)\,\frac{dV}{dt}=-I/CR \qquad \ldots\ldots(8.4)$$

$$\int_0^V Idt=\int_{I=0}^{I=-V/R}-CRdI=CV. \qquad \ldots\ldots(8.5)$$

This relation is confirmed by experiment, i.e. Q is the same whether it is estimated by charging or by discharging, and our assumptions are therefore justified. From these experiments we conclude that Q or ϕ_e is a significant magnitude, characteristic of the state of a capacitor, having a constant terminal-voltage, when its terminals are insulated.

We expect then that, when the terminals of the capacitor are insulated, and its configuration changes, Q will still remain constant. If so, in order to ascertain the behaviour of the electrometer when its terminals are insulated, we may substitute for V in (7.1) in terms of Q from (8.2), and perform the differentiation with respect to q subject to the constancy of Q. We thus have

$$F_q=-S\,.\,\frac{d}{dq}\left(\frac{1}{C}\right).Q^2. \qquad \ldots\ldots(8.6)$$

This law is confirmed by measurements; but, owing to the difficulty of maintaining the terminals truly insulated, it cannot be established as accurately as (7.1). Its truth is, however, a great part of the evidence for the propositions about Q.

§9. ENERGY IN A CAPACITOR

(8.6) is the law that would follow from general dynamical principles, if it were known that a pure insulated capacitor is a conservative system and that its potential energy is SQ^2/C. We therefore inquire whether there is any other support for these suggestions, taking into account equation (6) of Part I.

From (3.1)

$$\int_{t_1}^{t_2} IV\,dt = \tfrac{1}{2}C(V_2^2 - V_1^2), \qquad \ldots\ldots(9.1)$$

where V_1, V_2 are the voltages at t_1, t_2. It follows that, even if (6) were true of the instantaneous power converted into heat in a pure capacitor, the total energy and mean power dissipated in a complete cycle or in any number of cycles, would be zero. Experiment shows that it is zero or, more accurately, that any power dissipated in a capacitor in such circumstances can be associated with a departure of the capacitor from complete purity that can be detected in other ways. But it shows also that, in a pure capacitor, no energy is dissipated in even part of a cycle throughout which IV is of the same sign, although then, if (6) were true, power would be drawn from the generator and dissipated. Moreover, when a charged capacitor is disconnected from the generator by which it has been charged and is discharged through a resistor, V and I being observed during the discharge, power is dissipated in the resistor according to (6), being apparently drawn from the capacitor.

All this is consistent with the storage in the charged capacitor of potential energy W_Q, where, if the units of I, V are chosen to make the scale factor in (6) unity,

$$W_Q = \tfrac{1}{2}Q^2/C. \qquad \ldots\ldots(9.2)$$

Accordingly (8.6) will be identical with the law to be expected on general principles, so long as the units are chosen so that $S = \frac{1}{2}$. This is the reason for that choice in (7.1).

The potential energy can also be written

$$W_V = \tfrac{1}{2}CV^2. \qquad \ldots\ldots(9.3)$$

But it must *not* be concluded that the force tending to increase q, when V is constant, is $-dW_V/dq$. For, when V is constant, the capacitor alone is not a conservative system, because it can exchange energy with the generator by which V is maintained. The fact that F_q in these circumstances is given by (7.1) shows that the energy drawn from the generator in maintaining V is twice as great as, and opposite in sign to, the change in potential energy.

§10. GENERALIZATION OF CAPACITANCE AND CHARGE

The capacitors that we have considered so far ("simple" capacitors) are essentially pairs of conductors. Any conductor other than the pair that is relevant at all, in the sense that its removal would effect the capacitance of the pair, must be insulated from the pair, so that it is effectively part of the dielectric. But

sometimes we have to consider systems of more than two conductors, each pair of which is, or may be, maintained at a different non-zero voltage. The main, if not the only, examples of such systems offered in the older text-books are electrometers used heterostatically; but today more important and interesting examples are furnished by telephone cables and thermionic-valve circuits.

A study of the field equations suggests that the magnitudes *capacitance* and *charge* may have some relevance to such systems. Thus, suppose that we have N conductors, distinguished by suffixes $1, \ldots . m, \ldots . n, \ldots . N$, of which one is a screen surrounding all the rest; and that we consider the simple capacitor of which one plate is the conductor 1, characterized by v_1, and the other all the remaining conductors, characterized by v_0. If ϕ_m is the integral of (4.6) taken over m, and C_1 is the capacitance of this simple capacitor, then we must have

$$\phi_1 = -\Sigma \phi_m \qquad \ldots\ldots (10.1)$$

$$C_1 = \left| \frac{\phi_1}{v_1 - v_0} \right| = \Sigma \left| \frac{\phi_m}{v_1 - v_0} \right| . \qquad \ldots\ldots (10.2)$$

where the summation excludes 1; and, if we define C_{1m} by

$$C_{1m} = \left| \frac{\phi_m}{v_1 - v_0} \right| \qquad \ldots\ldots (10.3)$$

we must have

$$C_1 = \Sigma C_{1m}. \qquad \ldots\ldots (10.4)$$

Suppose now that we divide the conductors differently, taking 2 as one plate and all the others as the other plate. The capacitance C_2 of this simple capacitor will have components C_{2m} defined analogously to (10.3). By selecting in turn each of the conductors as one plate of the simple capacitor, we can find all the C_{mn}. It is now found to be a consequence of the field equations that

$$C_{mn} = C_{nm}. \qquad \ldots\ldots (10.5)$$

A more general consequence is that, if we assign to the conductors the values $v_1, \ldots . v_m, \ldots . v_n$, which may all be different, then, if ϕ_m and C_{mn} are defined as before,

$$\phi_m = \Sigma C_{mn}(v_m - v_n), \qquad \ldots\ldots (10.6)$$

$$\Sigma \phi_m = 0, \qquad \ldots\ldots (10.7)$$

where the summations extend over all the conductors. If we write $Q_m = \phi_m$, $v_m - v_n = V_{mn}$, these become

$$Q_m = \Sigma C_{mn} V_{mn}; \quad \Sigma Q_m = 0. \qquad \ldots\ldots (10.8)$$

This clearly suggests that the C_{mn} (which will be termed *mutual capacitances*) have the same physical significance as capacitances, and that, if the V_{mn} are identified with the voltages between mn, the Q_m have the physical significance of charges. Indeed, the system of conductors should be equivalent to a set of terminals $1, \ldots . m, n, \ldots .$, interconnected by simple closed capacitors of capacitance C_{mn}. The only difference between the system and a set of terminals interconnected by real simple closed capacitors is that each C_{mn} is determined, not by the configuration of two conductors only (namely the plate and its screen), but by the configuration of all the conductors.

Further, we should expect each Q_m to be equal to some integral $\int Idt$, where I is a current that flows to or from it during its charging or discharging. It is unnecessary for our purpose to define these relations generally; but it may be noted that Q_m should be $\int_0^\infty Idt$, where I is the current that flows in a circuit connecting the charged m to all the other conductors connected together. However, one matter requires more detailed attention. Hitherto, when Q has been the charge in a capacitor, it has been unnecessary to pay attention to the sign of Q, and the sign of I in the integral to which it is equal. Now Q_m is the charge on a single body, and, as (10.8) shows, it must sometimes be of one sign and sometimes of the other; it is important to lay down rules by which its sign can be ascertained. Formally, in order to solve all problems, these rules should be given in terms of the red and black terminals of ammeters and voltmeters discussed in Part I. But, since there is really no difficulty in the matter for anyone practised in electrical experiments, it suffices to say that the charge in a simple closed capacitor is to be regarded as positive if the voltage of the plate relative to the screen is positive, and that, when we write

$$Q=\int_0^\infty Idt, \qquad \ldots\ldots(10.9)$$

I must be the positive current flowing during discharge from the plate to the screen.

§11. PROPERTIES OF A COMPLEX CAPACITOR

All these propositions about a complex capacitor (for so we shall term a set of conductors such as is considered in the preceding section) are so far mere expectations, based on suggestions derived from the form of the field equations. There is a major difficulty in relating any of them to experiment. It is that, while the reasoning that leads to the propositions involves the assumption that the conductors are all isolated, any possible experiment requires leads between them. However, if we suppose that the presence of the leads does not destroy the equivalence between a complex capacitor and a set of terminals connected by simple closed capacitors, but merely modifies the mutual capacitances represented by the capacitances of those simple capacitors, then some experimental tests can be applied.

Thus, the propositions indicate that the formula corresponding to (9.3) should be

$$W_V=\tfrac{1}{2}\Sigma C_{mn}V_{mn}^2, \qquad \ldots\ldots(11.1)$$

where the summation is taken over all pairs without permutations. Since

$$V_{mn}=-V_{nm}; \quad V_{mo}=V_{mn}+V_{no} \qquad \ldots\ldots(11.2)$$

there are only $N-1$ independent V_{mn}; by solving the $N-1$ independent equations (10.8), we can obtain them as linear functions of the Q_m. Substituting these values in (11.1), we have

$$W_Q=\tfrac{1}{2}\Sigma p_{mn}Q_mQ_n, \qquad \ldots\ldots(11.3)$$

where the p_{mn} are functions of the C_{mn} subject to certain interrelations that need not be considered here. In an electrometer used heterostatically with insulated

terminals, F_q should be $-dW_Q/dq$; accordingly (11.3) predicts certain laws relating the forces on the electrometer to the charges on its parts, in which the constants are functions of the mutual capacitances. It is possible, as in many similar instances in various branches of physics, to apply tests that assume merely that these constants are indeed constants, and that do not require any knowledge of their actual values; if these tests are satisfied, some evidence will be obtained for the conclusions of this section without any detailed knowledge of the mutual capacitances.

Qualitative tests of this nature applied to electrometers are partially, but not completely, satisfied; it is almost certain that the discrepancies arise, not from any failure of the conclusions of this paragraph, but from the difficulty of taking into account *all* the mutual capacitances and of otherwise fulfilling exactly the conditions postulated. Perhaps better evidence of this nature is derived from valve-circuit theory, where again important conclusions can be reached, independent of accurate knowledge of the mutual capacitances. But the accuracy possible in all such tests is very low.

§ 12. MEASUREMENT OF MUTUAL CAPACITANCE: SCREENING

However, if a set of conductors is indeed equivalent to a network of closed capacitors, representing the mutual capacitances as modified by the leads, the capacitances of these capacitors can be measured by apparatus of the kind shown in figure 1.

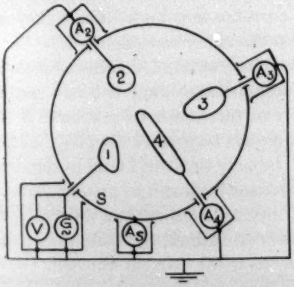


Figure 1. Scheme for the measurement of the mutual capacitance C_{mn} of any pair of conductors m, n.

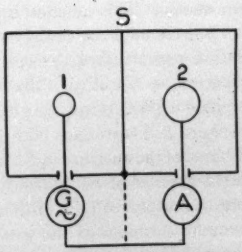


Figure 2.

Here 1, 2, 3, 4 are conductors within the screen S. G is a generator with one terminal connected to 1; A_2, A_3, A_4, A_5 are ammeters with one terminal connected respectively to 2, 3, 4, S; all the connecting leads are as thin as possible. The other terminals of the generator and ammeters are all connected together. The ammeters are all of low impedance, so that the voltage between 1 and all the other bodies is V, measurable by a voltmeter across the generator, and there is no voltage between any pair of these bodies.

If the generator and instruments had no effect beyond supplying and measuring current and voltage, it would be desirable to place them within S, in order that the

leads should be as short as possible. Examination of the equivalent network would then show that the only current I_m through A_m is the current through the capacitor C_{1m}; consequently we should have

$$I_m = C_{1m} \cdot dV/dt. \qquad \ldots\ldots(12.1)$$

Since I_m is measurable by A_m and V by the voltmeter, C_{1m} could be determined. (10.4) would be necessarily true in virtue of Kirchhoff's First Law and the method of measurement. C_{2m} etc. could be measured by substituting the other bodies in turn for 1; but experiment alone could determine whether $C_{mn} = C_{nm}$.

But the hypothesis is not true. The generator and instruments are really part of the system, having mutual capacitances with each other and with 1, 2, 3, 4, S. In order that (12.1) should be true in spite of these mutual capacitances, the generator and ammeters are placed outside S, and connected to the conductors within S by leads passing through apertures in S; and each is surrounded by its own screen connected to a common terminal (see § 14 below). This is a simple example of the art of screening, which forms an important part of A.C. technique. We shall not discuss it fully here; for one of us has discussed it in a recent publication (Hartshorn 1940). But some statement of the underlying principle is desirable

The principle involves more than that of the metallic cage mentioned in § 6. The object of the screens is not to isolate the generator and instruments completely from each other and from bodies inside S; but, firstly, to limit the extension in space of the electric fields of the bodies within the several screens and, secondly, to direct the capacitance currents associated with those fields to appropriate terminals of the equivalent network; or, in other words, not so much to abolish capacitance currents as to direct unwanted capacitance currents into paths where they are innocuous. This function can be explained with reference to figure 2.

The introduction of any conductor S between two others 1, 2, will decrease the mutual capacitance C_{12} to (say) C'_{12}; but it will also introduce two new mutual capacitances C_{1S}, C_{2S}. The total capacitance of the network between 1, 2 will become $C'_{12} + C_{1S}C_{2S}/(C_{1S} + C_{2S})$, and this will always be greater than C_{12}. If, as in figure 2, S surrounds both 1 and 2, C'_{12} will be very small and tend to zero with the area of the aperture in S; but the total capacitance will still be greater than C_{12} in the absence of S. Accordingly S in itself increases rather than decreases the mutual reaction of 1, 2. But if, as shown by the dotted lines, S is connected to the common terminal of the generator and ammeters, the current through C_{1S} is diverted from the ammeter to the dotted connection, and the current through C_{2S} is abolished, because there is no voltage driving it.

The action of the screens in figure 1 can be understood in the light of this simple example. But that example is merely an illustration of the principle involved; it must not be supposed that any feature of figure 1 is characteristic of all applications of the principle. Thus, it is not always desirable that each element should have its own screen, or that all screens should be connected to a common terminal. The application of the principle is so complex that it is scarcely possible to state any practical rule that is valid without limitation.

Experiments are not often made in exactly the way indicated in figure 1; but the various bridge and resonance methods for the measurement of mutual capacitances often (called "direct" or "partial" capacitances by communication engineers, who are chiefly concerned with them) are extensions of the same

principle, and depend upon (12.1) just as bridge methods for simple capacitors depend upon (3.1). These methods lead to consistent values in accordance with §§10, 11; in particular, such experiments fully establish the fact that C_{mn}, though dependent on the configuration of all the conductors, is not dependent on their voltages. The relation $C_{mn}=C_{nm}$ can also be established with high accuracy, provided that, in performing the experiment, the generator connected to m and the ammeter connected to n have their positions in the network interchanged without any disturbance of the leads that penetrate S. The values obtained often agree well with those calculated from the field equations. The measurements made by Rosa and Dorsey (1907) on the complex capacitor (a central cylinder with guard-ring extensions), used in their determination of $\mathbf{c}$, is an outstanding example. Such experiments provide the best evidence for the general theory of complex capacitors, i.e. in traditional language, "systems of conductors".

Finally, precision methods of measuring the capacitance of standard capacitors are essentially methods of measuring mutual capacitance. It is therefore impossible in principle that they should lead to entirely definite values. For, as we have seen, the mutual capacitance of two conductors depends on all the other conductors with which it forms a definite group in the sense of §6; moreover, measurable mutual capacitances are always modified by the presence of leads; the very purpose of a standard capacitor forbids that it should always be associated with the same conductors or joined to the same leads. However, the margin of uncertainty can be made very small, e.g. 2 parts in a hundred thousand for a capacitance exceeding $0{\cdot}1\,\mu$F.; and, as indicated in §6, the rules of addition in parallel are satisfied by some values within this margin.

In some exceptional circumstances it is possible to assign values to elastances within a comparable margin of uncertainty, and to show that elastance is additive in series. But it will not usually be possible to prove this with high accuracy by the use of standard capacitors; for, as usually constructed, they do not satisfy sufficiently nearly the condition, stated in §6, for the ideal addition of elastance.

§13. ABSOLUTE CHARGE

In the foregoing discussion the charge on a body has appeared as a property that it possesses in virtue of its relation to other bodies with which it forms a capacitor. On the other hand, in the classical theory of electrostatics, charge appears as an "absolute" property of a body, independent of its relation to any other body, and moreover, as one that can be measured without the knowledge of any of the algebraic laws involving charge, and, in particular, of those algebraic laws that concern capacitors. We have to inquire on what facts such a conception of absolute charge is, or could be, based.

Suppose that a closed capacitor is charged by connecting the plate to the screen through a battery, and that the plate is then discharged by connecting it to the screen through an ammeter. If I is the current through the ammeter during the discharge, the charge acquired by the plate during the charging is given by $Q=\int_0^\infty I\,dt$. Suppose now that the plate, after having been charged, is removed from the screen in which it was charged and, remaining always insulated, is transferred to the interior of another screen. If it is now connected to this screen

through an ammeter, $\int_0^\infty Idt$ should have the same value as before, whatever the form of the second screen, wherever the plate is placed within this screen, whether the first screen is insulated from the second, or whether a voltage, zero or non-zero, is maintained between them by a conductor or battery. In classical electrostatics this proposition is stated as a consequence of the theorem that the charge on a conductor resides wholly on the outside of it; but it does not follow from anything in our previous discussion. If it is true, it must be established by direct experiment. If it proves to be the fact that, whenever a body, charged as part of a capacitor, simple or complex, is discharged to an enclosing shield, $\int_0^\infty Idt$ turns out to be the same as the charge that it acquired as part of the capacitor, then *charge* will have a significance independent of any particular capacitor, though still not independent of the general properties of capacitors.

The experiment would be difficult to perform with any accuracy, and the test has probably never been applied. But an equivalent test may be devised by noting that the law $Q=CV$ is inseparable, according to our exposition, from $Q=\int_0^\infty Idt$. If, instead of discharging the body to the second screen, we measure the capacitance of the closed capacitor, formed by the body and the second screen, and also the voltage across the capacitor, then we should find that, though C and V depend on the form of the second screen and on the position of the charged body within it, CV is independent of both and equal to the charge that the body acquires in the first screen.

However, the test is still not convenient, since it requires the measurement of a different capacitance each time a charge is to be measured. The celebrated "Faraday cage" is essentially a device for overcoming this difficulty. It is illustrated in figure 3, which shows diagrammatically the complete electrical system involved in any attempt to perform the experiment accurately. Here the internal screen 1 is the "cage" into which the charged body 3 is introduced, so that 1 and 3 form a closed capacitor of which the product VC is required. An outer screen 2 surrounds 1, and a voltmeter V_{12} is connected between them; this arrangement provides another closed capacitor whose capacitance C_{12} and voltage V_{12} are accurately measurable. Now the complete capacitor 3, 1, 2 has that special form for which elastance is strictly additive (§6), and which is therefore completely represented by two simple capacitors in series, C_{13} and C_{12} in figure 3. Thus, when this complex capacitor is charged, which is the case when the body 3 enters the cage 1, each of these simple capacitors receives the same charge. Consequently, if V_{12} is originally zero, the value that it will acquire when the body 3 bearing charge Q enters the cage 1 is given by

$$Q=C_{12}V_{12}, \qquad \ldots\ldots(13.1)$$

where C_{12} includes the capacitance of the voltmeter. Accordingly a knowledge of C_{12} alone is required, and this is independent of the form of 3 and of its position within the cage.

It is now quite practicable to apply the test. 3 is given a known charge Q by charging a closed capacitor, of which 3 forms the plate and 4 the screen, to a

voltage V_{34} by a battery between 3 and 4. 3 is now introduced into 1. If V_{12} were originally zero, it should now be given by

$$C_{12}V_{12} = C_{34}V_{34}. \qquad \ldots\ldots(13.2)$$

If this relation is always fulfilled, we have a way of measuring the charge on a body independently of the capacitor in which it received its charge.

V_{12} need not be zero originally. If it were not, then (13.1), (13.2) should be true, if δV_{12} is substituted for V_{12}, where δV_{12} is the change of V_{12} when 3 is introduced. This observation is important, because it leads to a method of measuring Q that does not require the prior establishment of methods of measuring C and V, or indeed of any other magnitude. Suppose that the voltmeters V_{12}, V_{34} are originally uncalibrated. We make some mark on the scale of V_{34}, and assert that, when its pointer stands at this mark, we are going to say that 3, being within 4, has the charge 1. We transfer 3 to within the cage, and, V_{12} having been originally zero, note the reading of V_{12}, and call this 1. We discharge 3 to the cage (by virtue

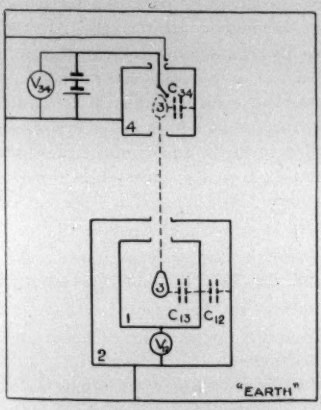


Figure 3. The "Faraday Cage" Experiment.

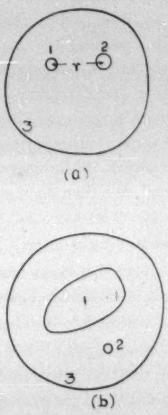


Figure 4. The problem of three charged bodies. Coulomb's Law is an approximation for case (*a*) and $F_r = -Q_2 \frac{dv}{d_r}$ an approximation for case (*b*).

of principles already laid down, this should not change the reading of V_{12} and experiment confirms this), put 3 back into 4, charge it so that again it has charge 1, and transfer it again to the cage. The reading of V_{12} now we mark 2. Repeating the process, we get points on the scale of V_{12} marked 3, 4, ... In order to ascertain whether these marks on V_{12} provide a consistent scale of charge, we change the voltage V_{34} or the body 3, or both, until we find that, when 3 charged in 4 is transferred to 1, it causes V_{12} to change from 0 to 2. Then, repeating this process, we find that the second transfer changes V_{12} from 2 to 4. If the rules of addition, tested by a sufficient variety of such experiments, are always fulfilled, the calibrated

scale of V_{12} will measure charge, and we shall have established charge as an independently measurable magnitude; we shall have gone a long way to providing evidence for the fundamental postulate of classical electrostatics. Of course nowadays this independent measurement of Q is of little importance; it is easier and more accurate to measure Q as the derived magnitude CV. But it is remarkable that no experiment of this kind, so necessary as a starting point for classical theory, appears in either the standard treatises or the histories. However, this becomes less surprising when we observe that all the laws of physical addition, even those of length and mass, have almost always been accepted purely intuitively. They are so fundamental that the underlying assumptions are made unconsciously. It is, however, desirable that they should be brought to light.

But, though we have now dissociated Q from any magnitude characteristic of capacitors, we have not rendered it entirely independent of them. It still is significant only in connection with the charging of capacitors. Two further experiments are needed to establish complete independence. One arises from the discovery that bodies that have not been charged as part of a capacitor may produce a reading of V_{12} when they are introduced into the cage. In order to establish that the magnitude charge may fitly characterize such bodies, we must show that the introduction into the cage of n bodies, each of which when introduced singly gives the same value of δV_{12}, gives a reading of $n\delta V_{12}$.

For the second experiment, we require two similar bodies 3 as shown in figure 3, and we break the connections of 2 and 4 to earth, and connect these screens together. We then have in figure 4 a complex capacitor consisting of three simple closed capacitors in series, namely those shown dotted in figure 3. We connect the two bodies 3 by a battery in series with an ammeter, and measure $\int Idt$, where I is the current that flows during the charging of the complex capacitance. We now remove 3 from within 1, and introduce into the cage 1 the body 3 which has been charged within 4. We shall find that the charge on 3 so measured is $\int Idt$ measured during the charging. We have then proved that the transport of a charged body between the two positions 3 in figure 2 has precisely the same effect, in respect of this apparatus, as the passage of a current I between those two positions, so long as the charge on the body is $\int Idt$, and *vice versa*. This leads us to inquire whether the transport of a charge is equivalent to the passage of a current in respect of other apparatus, e.g. that for measuring the magnetic field produced by a current. We find that it is, so long as the path traversed by the charge during its transport is the same as the path of the current. When we have proved that, we have proved that charge has a significance wider than that of electrostatic experiments, all of which involve capacitors; we shall have proved, in the language of our ancestors, the "identity of all forms of electricity".

§ 14. "EARTH"

Earth is a conception hardly less fundamental in classical theory than *charge*; but the explanations offered of what constitutes earth* are usually very vague, and

* We leave on one side problems concerning the "atmospheric gradient" or old-fashioned telegraphy with an "earth return". For in them "earth" undoubtedly means The Earth. But it is perhaps worth noting that there was a period (roughly that of the 1870's) when electrical nomenclature was becoming standardized and telegraphy was the only form of electrical engineering. The present uses of "earth" are probably derived in some measure from "earth returns".

sometimes demonstrably false. In the absence of a formal definition the meaning of earth can be ascertained only by examining the statements that are made about it.

Thus *earth* is the thing to which zero potential is assigned in conventional discussions of "systems of conductors". Since differences of potential are voltages, it is the thing from which voltages are measured, i.e. the "zero plate" of §5. Now if there is a terminal to which many more conductors are connected than to any other, it is clearly convenient to take that conductor as the zero plate. In apparatus containing many screens through which leads pass, and which are therefore connected together, such as that of figure 1, the terminal to which the screens are connected is such a terminal. If therefore *earth* means the common terminal of screens in such apparatus, it has a clear significance in the light of our exposition, which is all based on the conception of capacitance. Moreover, such a use of the term would agree with ordinary laboratory usage; for such a common terminal is called "earth" in that usage; indeed, the avoidance of *earth* in our discussion of screening must have appeared artificial to many readers. Further, the usage has strong justification; for in such apparatus the sheathing of cables almost always forms part of the screens, and this sheathing is always connected to The Earth.

If, however, the observer and all his apparatus are enclosed in a complete screen through which no leads pass, this screen will always be used as the common terminal of less perfect screens within it and, so far as is possible, as one terminal of generators and detectors. There is then no need that the screen enclosing the entire apparatus should be connected to The Earth; indeed, in a high-voltage laboratory that may be positively undesirable or impossible; one man's earth may have to be another man's H.T. terminal. Accordingly, even if *earth* could always be identified with a common terminal in the sense just described, it would still be necessary to recognize that "earth" does not always mean The Earth, and, more generally, that there is no single body that can always be taken as earth.

There is another statement about "earth" made or implied in conventional text-books, especially in the description of the classical experiments of electrostatics, on which the whole subject is supposed to be based. It is that, if any charged body is connected to earth, it will be wholly discharged. According to §13, this will be true only if "earth" means a screen surrounding the body completely or practically completely. If that is what "earth" means, earth cannot possible be a single body everywhere at the same potential; for, if there are two charged capacitors in series, the screen of one is necessarily the plate of the other. Nevertheless, the use of "earth" in this sense may have some significance in connection with the said classical experiments. For if they were conducted in the open, some of them would fail owing to the Earth's atmospheric gradient; if they are to succeed, they must be conducted within a screen, which will have to be the more complete, the more accurate the real experiments. The classical experiments are so inaccurate that the very incomplete screen formed by the partially conducting walls, ceiling and floor of the building in which they are conducted may suffice. When a charged body is discharged by touching it with the finger or with a wire connected to the water pipes, that may be because such contact provides a sufficiently conducting path between the body and the imperfect screen by which it is surrounded.

But another explanation is possible. The charge in a capacitor will be greatly reduced if its terminals are connected to an uncharged capacitor of much greater capacitance. It is possible that a charged body is discharged by the finger of the observer, not merely because he provides a conducting path to the screen, but because his body forms with all those conductors, with which the charged body has mutual capacitance, a capacitor of much greater capacitance; or, more simply, because his body is much larger than the body he discharges.

This, we suggest, is the basis of the curious statement to be found in some well known text-books (it appears to have grown commoner during the last generation) that The Earth acts as an "earth", because its capacitance is much greater than that of any body concerned in terrestrial experiments. But, though it may provide the basis, it provides no justification. For, firstly, the statement is not actually true; the capacitance of The Earth, as ordinarily calculated, is no greater than that of certain electrolytic condensers that are often used in laboratory experiments. Secondly, and more important, the capacitance of The Earth cannot possibly have any relevance to any terrestrial experiment. The facts that give significance to the conception of the capacitance of a single body (not of a pair of associated bodies) are these: (they are predicted by the field equations and confirmed by experiment.) If in a closed capacitor the distance between the plate and the screen is increased without limit, the capacitance of the capacitor tends to a lower limit; the capacitance of a single body means this lower limit towards which the capacitance of a closed capacitor tends. But since there is no accessible conductor whose distance from The Earth is large compared with The Earth's dimensions, this limit can have no relevance to experiments conducted on The Earth.

§ 15. POINT CHARGES

Point charges play a very important part in classical theory—much more important than that which they play in experimental physics. It seems therefore important to show that the propositions that we have set forth lead to the chief propositions about points charges. (These are Coulomb's law, and

$$F = e\mathscr{E} = e \,.\, dv/dq, \qquad \ldots\ldots(15.1)$$

where F is the force in the direction of the coordinate of q acting on a point charge e in a field of potential v (§ 5)).

Let us consider then the force between two charged conductors, which should lead us to Coulomb's Law. For reasons already given we must include in our system an earth, i.e. a third conductor surrounding the two small ones (figure 4*a*); but in order to simplify the problem we shall assume that the third conductor is so large that the distance r between 1 and 2 is very small compared with their distances from 3, but large compared with the linear dimensions of 1 and 2. Unless the charges on 1 and 2 are equal and opposite, there must be a charge equal to their difference on the inner surface of 3; there may also be charges on its outer surface, but they do not concern us. We have from (11.1) with an obvious notation

$$W_V = \tfrac{1}{2}(C_{12}V_{12}^2 + C_{23}V_{23}^2 + C_{13}V_{31}^2) \qquad \ldots\ldots(15.2)$$

with

$$Q_1 = C_{12}V_{12} + C_{13}V_{13}, \qquad \ldots\ldots(15.3)$$

$$Q_2 = C_{12}V_{21} + C_{23}V_{23}, \qquad \ldots\ldots(15.4)$$

$$V_{23} = V_{21} + V_{13}. \qquad \ldots\ldots(15.5)$$

Substituting for the Vs in (15.2) from (15.3, 4, 5), we find after some reduction

$$2W_Q(C_{12}C_{23}+C_{23}C_{13}+C_{13}C_{12})=Q_1^2(C_{12}+C_{23})+Q_2^2(C_{12}+C_{13})+2Q_1Q_2C_{12}. \quad \ldots\ldots(15.6)$$

In case (*a*), if the bodies 1 and 2 are small compared with their distances from the screen and from each other, C_{12} is small compared with each of C_{13}, C_{23}. It follows then from (15.6) that, if F_r is the force tending to increase the distance r between 1, 2, then

$$F_r=-dW_Q/dr=-\left\{\tfrac{1}{2}Q_1^2\frac{d}{dr}\left(\frac{1}{C_{13}}\right)+\tfrac{1}{2}Q_2^2\frac{d}{dr}\left(\frac{1}{C_{23}}\right)+Q_1Q_2\frac{d}{dr}\left(\frac{C_{12}}{C_{13}C_{23}}\right)\right\}. \quad \ldots\ldots(15.7)$$

Now, according to a well known result (see e.g. Jeans, *Electricity and Magnetism*, §116), if a, b are the radii of spheres to which the bodies 1, 2 approximate, and if $ab \ll r^2$

$$C_{13}=\kappa a; \quad C_{23}=\kappa b; \quad C_{12}=\kappa ab/r, \quad \ldots\ldots(15.8)$$

where κ is the permittivity of the medium surrounding the bodies. Hence the terms of (15.7) in Q_1^2 and Q_2^2 vanish, and we have

$$F_r=-Q_1Q_2\,.\,d/dr(\kappa ab/r\,.\,1/\kappa^2 ab) \quad \ldots\ldots(15.9)$$

$$=+Q_1Q_2/\kappa r^2, \quad \ldots\ldots(15.10)$$

which is Coulomb's Law. Thus, although the law is not strictly true for small charged conductors, as is obvious from (15.7), it is an approximation that becomes true in the limiting case in which the charged bodies become infinitely small.

Consider now the force on a single small charged body 2 in the field between two large bodies 1 and 3 (figure 4*b*). This is another case of the general problem of three charged bodies, and (15.6) again holds; but now C_{12}, C_{23} are both smal compared with C_{13}, as in (say) the oil-drop experiment. So we have

$$F_q=-dW_Q/dq=-\tfrac{1}{2}Q_1^2\frac{d}{dq}\left(\frac{1}{C_{13}}\right)-\tfrac{1}{2}Q_2^2\frac{d}{dq}\left(\frac{1}{C_{12}+C_{23}}\right)$$

$$-Q_1Q_2\frac{d}{dq}\left(\frac{1}{C_{13}}\cdot\frac{C_{12}}{C_{12}+C_{23}}\right), \quad \ldots\ldots(15.12)$$

where q determines the position of the small conductor relative to the large ones, whose relative position is constant.

The term in Q_1^2 represents a force, important in some circumstances, tending to bring 2, even if it is uncharged, into the region where the field due to the charges on 1 and 3 is strongest. Since e in (15.1) is our Q_2, there is no corresponding term in (15.1) independent of Q_2. The term in Q_2^2 represents (in conventional language) the "image force" between Q_2 and the charges that it "induces" in the large conductors. If v is independent of Q_2, as the form of (15.1) suggests, there is again no corresponding term in (15.1); but one can be introduced by including in v a part due to the "induced charges"; a part due to the charge Q_2 on 2 must not be introduced, for such a part would not give rise to a force on 2. Accordingly (15.1) in its natural interpretation is at best an approximation valid only when the terms of (15.7) in Q_1^2 and Q_2^2 are negligible. That in Q_1^2 is negligible, if the field

due to the charges on 1 and 3 is sufficiently uniform and the dimensions of 2 sufficiently small. The term in Q_2^2 is negligible, if Q_2/Q_1 is sufficiently small and if also 2 is sufficiently far from 1 and 3; for, when it is far from 1 and 3, dC_{12}/dq and dC_{23}/dq tend to be equal and opposite.

If these conditions are fulfilled (15.7) reduces to

$$F_q = -Q_1 Q_2 \cdot \frac{d}{dq}\left(\frac{1}{C_{13}} \cdot \frac{C_{12}}{C_{12}+C_{23}}\right); \qquad \ldots\ldots(15.13)$$

and this is the equation that really has to be compared with (15.1). In that equation v is the potential relative to the screen 3 of the point occupied by 2, when 1 is at its actual potential V_{13}. This may be identified with the voltage that would be established between 3 and a small uncharged conductor in the position of 2, if the voltage between 1 and 3 were increased from 0 to V_{13}. The small conductor, being insulated, cannot become charged during the process; and we have from (15.4)

$$0 = C_{12}V_{21} + C_{23}V_{23} = C_{12}(v - V_{13}) + C_{23}v$$

or

$$v = V_{13} \cdot C_{12}/(C_{12}+V_{23}). \qquad \ldots\ldots(15.14)$$

Hence (15.13) becomes

$$F_q = -Q_1 Q_2 \cdot \frac{d}{dq}\left(\frac{v}{C_{13}V_{13}}\right) \qquad \ldots\ldots(15.15)$$

or from (15.3),

$$= -Q_2 \cdot \frac{d}{dq}\left(v \cdot \frac{Q_1}{Q_1 - C_{12}V_{12}}\right), \qquad \ldots\ldots(15.16)$$

(15.16) is identical with (15.1) only if $C_{12}V_{12} \ll Q_1$. This does not follow immediately from (15.3), (15.4) together with $Q_2 \ll Q_1$ and $(C_{12}, C_{23}) \ll C_{13}$; for the two terms in (15.4) may be of opposite sign. But it can hardly be doubted that there are conditions in which (15.16) approaches (15.1) as a limit as Q_2 approaches zero. We leave to the mathematician the task of deciding exactly what those conditions are; they will certainly involve the interpretation given to v in (15.1). Further, a correction, which will certainly involve terms in Q_2/Q_1 only, has to be made for the fact that, while (15.1) relates to a state in which V_{13} remains constant during the motion of Q_2, we have calculated the force on the assumption that Q_1 remains constant during that motion. It is sufficient for our purpose to note that (15.1) can be deduced by the argument that we have presented only as a highly specialized limit.

The complexity of the argument necessary to deduce the very simple classical formula (15.1) may suggest the criticism that, though our exposition of electrostatics may be ultimately sounder than that of classical theory (at least if that theory is supposed to have any relation to experiment), the classical treatment is always to be preferred in dealing with any practical problem. We reject any such suggestion. The physicist who applies the simple formula (15.1) to any real experimental system must always make due allowance for the difference between his actual system and the ideal system for which the formula is strictly valid. The simplicity of (15.1) is obtained only at the cost of complexity of interpretation in terms of measurable quantities, and our methods remove this complexity. For general problems in experimental physics we suggest that our methods are far more perspicuous physically, and no more difficult algebraically, than classical

methods. This is particularly true of the problems (§§ 10–12) arising in "systems of conductors"; actual experience has convinced us that the treatment of such problems in terms of the equivalent capacitance network has great advantages.

§ 16. CHARGES ON INSULATORS

It has been assumed in the two preceding sections, explicitly or tacitly, that charged bodies are always conductors. This is not true. Of the bodies which, when introduced into a Faraday cage produce a non-zero δV_{12}, some are insulators. Such bodies cannot form a plate of a capacitor, cannot be charged or discharged by a current circuit, and have no measurable voltage to other bodies; accordingly many of the propositions of those sections have no application to them. Nevertheless they have a determinate charge, because they satisfy the criterion of § 14, namely that n equivalent bodies, introduced together into the cage, produce n times the effect of one of them. Further, if we suppose that a conductor and an insulator, having the same charge according to the Faraday cage, differ only in the distribution of the charge, all propositions concerning point charges must be equally true of both of them; for in the limit when the dimensions of the body are zero, the distribution of the charge in the body cannot matter. Accordingly (15.1) can be applied to the measurement of the charge on an oil drop, by observing the force on it in a uniform field, although the drop may be an insulator; or to the deflection of a cathode ray in an electric field, without making an assumption about the constitution of an electron, except that it is small and charged.

But the distribution of charge on an insulator enters into other problems. They can be solved in a manner that we shall not discuss in detail by means of a proposition about the relation between the field v (§ 6) about an insulator and the distribution of charge in and over it. The conventional proposition is that if σ is the surface density of the charge and ρ the volume density, then v must satisfy the differential equations

$$4\pi\sigma = -\left(\frac{dv}{dn_1} + \frac{dv}{dn_2}\right), \qquad \ldots\ldots (16.1)$$

where n_1, n_2 are the normals to the surface on opposite sides of it

$$4\pi\rho = -\left\{\frac{d}{dx}\left(\kappa\frac{dv}{dx}\right) + \frac{d}{dy}\left(\kappa\frac{dv}{dy}\right) + \frac{d}{dz}\left(\kappa\frac{dv}{dz}\right)\right\}. \qquad \ldots\ldots (16.2)$$

The direct evidence for these propositions is very slight. Some evidence in favour of (16.1) can be obtained by means of "proof planes", i.e. small conductors applied to the surface of the insulator and then transferred to a Faraday cage, if it is assumed that the proof plane takes up the surface density of the surface to which it is applied. (It may be noted incidentally that, by means of proof planes, evidence can be obtained for the presence of the charges $-Q_1$ and $+Q_1$ on the inner and outer surfaces of the intermediate plate of the two capacitors in series discussed in § 5.) Volume charge can be detected only by breaking the body and transferring the fragment to the cage. Actually volume charges in solids are at present of little experimental importance. The force between two bodies, one at least of which is an insulator, arising from their charges, can be obtained by means of a proposition

that might have been stated earlier. It can be shown by purely mathematical reasoning that (11.1), (11.2) are equivalent to

$$W = \frac{1}{8\pi}\iiint \kappa \left\{\left(\frac{dv}{dx}\right)^2 + \left(\frac{dv}{dy}\right)^2 + \left(\frac{dv}{dz}\right)^2\right\} dxdydz. \quad \ldots\ldots (16.3)$$

It is assumed that this formula for the energy holds even when some or all of the charges reside on insulators. The results deduced from this assumption, together with (16.1), (16.2), apparently agree with such experiments as can be made.

Finally, it might appear that, if the dielectric forming part of a capacitor were charged, the rule for calculating capacitance given in §4 would no longer hold. Experiment, in accordance with deduction from (16.1), (16.2), shows that such charges, though they modify the values of v, do not modify the capacitance.

§17. INDUCED CHARGE

In the conventional exposition of electrostatics a magnitude called *induced charge* is introduced, and considerable space is devoted to a discussion how far "induced" and "real" charges have the same properties. The expression *induced charge* is often useful in drawing attention to the fact that, when an insulated conductor is introduced into an electric field, the distribution of the charge on it, as measured by ϕ_e, changes, while its total charge, measured by $\int Idt$, does not. But it appears to us that attempts to give the expression definite quantitative significance introduce confusion rather than clarity.

REFERENCES.

CAMPBELL, N. R. and HARTSHORN, L., 1946, *Proc. Phys. Soc.*, **58**, 634.
HARTSHORN, L., 1940, *Radio-Frequency Measurements by Bridge and Resonance Methods* (London : Chapman and Hall).
ROSA and DORSEY, 1907, *Bull. Bur. Stand.*, **3**, 43.

Measurement of Absolute Humidity in Extremely Dry Air

BY A. W. BREWER, B. CWILONG AND G. M. B. DOBSON,
Oxford

MS. received 29 *May* 1947

ABSTRACT. To meet an urgent demand for a hygrometer capable of use at all heights in the atmosphere, the dew-point hygrometer, which is shown to have many advantages, has been developed primarily for use in aircraft to measure dew-points, or rather frost-points, down to −90° C.

It is necessary that the instrument should operate at the lowest possible frost-points as it has been discovered that the air in the stratosphere is very dry. Laboratory studies of the deposition of water and ice from the vapour at low temperature are described.

Below −90° C. it is not possible to operate a frost-point hygrometer because the deposit is in the form of an invisible glassy layer, but the instrument gives correct results at temperatures close to this limit.

Details are given of the construction of different forms of hand-operated hygrometers, and work is now going on to develop a fully automatic frost-point hygrometer.

§1. INTRODUCTION

METEOROLOGISTS have long felt the need for an instrument capable of measuring the water content of the upper air at altitudes where the low temperatures preclude the use of the usual methods. The method of measuring the water content of air, or any other gas, down to a concentration of 0·4 mg./m^3 which is described here, has been developed primarily for use in the upper air from aircraft, but it is quite suitable for experimental use for any other purpose.

The interest in the humidity of the upper air is due to the fact that the water content at any level usually gives an indication of the past history of the air at that level. For this particular problem a hygrometer is needed which retains its accuracy at very low relative humidities, for instance one which can distinguish between 1% R.H. and 2% R.H. in air at a very low temperature. To do this the hygrometer must be capable of measuring small quantities of water vapour with an accuracy not previously attainable.

The average temperature of the stratosphere in temperate latitudes is about $-56°$ c. (This is the temperature of the stratosphere in the ICAN standard atmosphere), and air saturated at this temperature would have a water vapour content of about 18 mg./m^3. When this work was begun it was expected that this concentration, or a value only slightly below it, would be the lowest that would have to be measured. When, however, flights were made into the stratosphere it was found that the relative humidity there was of the order of only 1 or 2%, corresponding to saturation temperatures down to $-80°$ c. In order, therefore, to measure the water content of the stratosphere it has been necessary to develop a hygrometer capable of measuring water contents of less than 1 mg./m^3.

Most of the hygrometers ordinarily used in meteorological work are quite unsuitable for use at low water contents and low temperatures. At low temperatures the usual wet and dry bulb psychrometer has a vanishingly small temperature difference even for 0% R.H., and also the constant in the psychrometric formula is in doubt (Griffiths and Awbery 1935). The hair hygrometer, the gold-beater's skin hygrometer, and other hygrometers employing organic elements sensitive to relative humidity have a large lag which increases rapidly at low temperatures so that they cannot be used. Electrical conduction hygrometers, such as the Gregory humidiometer, have a large temperature coefficient, which seriously interferes with their use at low temperatures, while in addition the lag increases to an undesirably large value. For further discussion of these difficulties, reference may be made to Ebert (1937) and Simons (1936), who discuss the difficulties of low-temperature hygrometry. Glückauf (1945) discusses the effect of lag on the hair hygrometers in meteorograph measurements and Diamond *et al.* (1940) discuss the properties at low temperatures of the electrical hygrometric element of the U.S. radiosonde. Dymond (1947) gives a similar discussion for the gold-beater-skin element of the British, Kew pattern, radiosonde; he also quotes results obtained by Harrison and Brewer (1944) of comparison ascents in which an aircraft, equipped with a hygrometer of the type described in §3 of this paper, circled an ascending radiosonde balloon.

The only methods which appear at all practicable are the dew (or rather hoar frost) point hygrometer, and a radiation absorption method. The latter method could only be operated in the 6·7 μ absorption band on account of the

very small quantities of water involved and, even so, would require very long absorption paths.

The optical absorption hygrometer was carefully considered, but it was decided that it would present difficulties at very low vapour pressures and also much basic work on the absorption by water vapour of infra-red radiation at low pressures and temperatures would be required. The radiation method is most likely to be successful if the absorption of solar radiation in the 6·7 μ band is measured, so as to give the total water absorption above the observer.

The dew point hygrometer on the other hand has been used by several workers down to about $-30°$ C. without giving special difficulties and it seemed hopeful that it could be developed to operate at much lower temperatures. The instruments described here have been used regularly at frost-points down to $-80°$ C., but there is a natural limit to their operation at about $-90°$ C. which will be discussed fully in § 2. Since throughout this work the water contents measured have almost invariably been below the saturation vapour pressure at 0° C., the instruments are usually referred to as frost-point hygrometers, particularly as they are more sensitive to a hoar-frost deposit than a dew deposit, and this name will be used below.

The frost-point hygrometer has many advantages. The water content is measured upon a nearly logarithmic scale quite independent of the actual air temperature, and this makes for high discrimination at low relative humidities, provided that the water content is not so very low as to correspond to a frost-point below $-90°$ C. For meteorological purposes also the observations can be used directly by considering the frost point, usually plotted with the air temperature, upon a thermodynamic diagram. For use in an aircraft, it is also very helpful to be able to bring the air into the aircraft in a duct and measure the frost-point in a convenient apparatus, for, provided that the walls of the duct do not give off or absorb water vapour, then there is no change in the dew- or frost-point. A hygrometer which measures the relative humidity has, by comparison, many difficulties. Another very real practical advantage is its basic stability. For calibration it is only necessary to calibrate the thermometer which measures the thimble temperature. The thermometry involved is very simple, and re-calibration is not normally required even after long periods of either use or disuse. Another advantage is that supersaturation which is not uncommon in the upper air is shown quite simply, as the frost-point is then measured as being higher than the air temperature, which is measured separately.

To illustrate the practical use of the frost-point hygrometers described below the results of two ascents are shown in figure 1 and figure 2. Figure 1 is an ascent to 12 km. made during unsettled weather and shows moist air in the troposphere and the usual very dry air in the stratosphere. Figure 2 is an ascent to 4 km. made in an anticyclone with observations at frequent intervals to show the details of the distribution of water vapour in the stable layer between 1 and 2 km. Both these ascents were made with eye-observation instruments of the type described in § 3. It is hoped to discuss the meteorological aspects of these and other ascents elsewhere.

§ 2. DEPOSITION OF WATER VAPOUR AT VERY LOW TEMPERATURES

In addition to the obvious difficulties involved in observing very small amounts of dew or frost, a number of practical difficulties are found when the

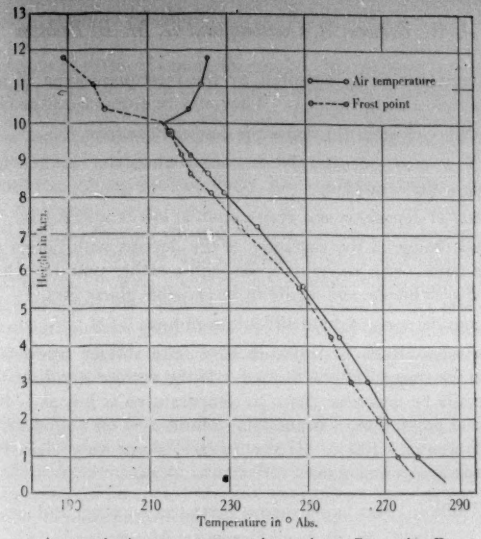


Figure 1. Frost-points and air temperatures observed on Boscombe Down, 30 May 1945. 13.00 G.M.T. The frost-point at the top is the lowest yet observed.
(Reproduced, by kind permission, from the ' Proceedings of the Royal Society '.)

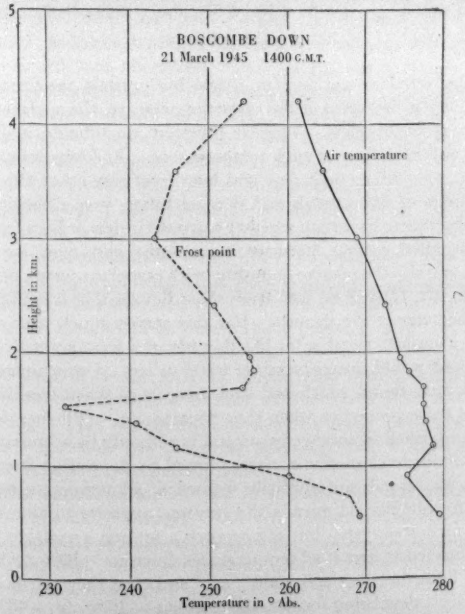


Figure 2. Frost-points and air temperatures observed in an anticyclonic inversion. 21 March 1945.
(Reproduced, by kind permission, from the ' Proceedings of the Royal Society '.)

humidity of very dry air is measured by the frost-point type of hygrometer. Some of these have a wider interest. They may be summarized as follows:—

(*a*) Uncertainty whether the deposit is dew or hoar frost.

(*b*) Hoar frost is not necessarily deposited when the thimble temperature is below the frost-point.

(*c*) Slowness of deposition and evaporation at low temperatures.

(*d*) Gradual change in the character of the deposit with falling frost-point, from large white ice crystals, through a rather uniform white deposit, to a bluish bloom, and finally to an invisible glassy film.

We shall consider each of these difficulties in turn.

(*a*) When water vapour is deposited on a solid surface below 0° c. it does not follow that ice crystals will be formed. If the surface is polished and clean, dew will generally be deposited down to temperatures as low as −40° c., while by taking special precautions it is possible to form dew on a solid surface at any temperature down to −100° c. If the surface is not very clean ice may be formed at temperatures well above −40° c.

Table 1. Saturation vapour pressures on water and ice
(Brunt, *Physical and Dynamical Meteorology*)

Temperature (°K.)		270	260	250	240	230	220	210	200
Sat. v.p. (mb.)	over water	4·89	2·25	0·97	0·38	0·14	0·045	0·013	0·0035
	over ice	4·76	1·98	0·77	0·27	0·09	0·027	0·007	0·0017
Ratio $\frac{\text{v.p. ice}}{\text{v.p. water}} \times 100$		97	88	79	71	65	60	54	49

If ice is deposited at, say, −5° c., a few big crystals are formed. On the other hand if dew is deposited at the same temperature, the surface is uniformly covered with very fine droplets. There is, therefore, no difficulty in distinguishing between dew and hoar frost at such temperatures. At lower temperatures, say about −30° c., to −40° c., both dew and hoar frost may cover the surface with small water drops or fine crystals and it is, at times, very difficult even for an experienced observer to be certain whether a deposit is dew or hoar frost. Table 1 shows the saturated vapour pressure over water and over ice at different temperatures, and also the relative humidity, with respect to water, of air saturated with respect to ice. It will be seen from these figures that it is important to be sure of the character of the deposit. For this reason much time was spent in trying to find a surface, suitable for the thimble of a frost-point hygrometer, on which the deposit would always be either water or ice. Unfortunately, although clean surfaces were found which had such properties when freshly made, they tended to lose these properties when they became old. (Further details of this work will be published in another paper and it will only be summarized below.)

On a freshly made clean zinc surface, formed by deposition from the vapour in a vacuum, ice crystals are normally formed at all temperatures below 0° c. A surface of freshly cleaved mica is the extreme opposite to zinc and it seems impossible to form ice crystals initially on such a surface, a uniform film of water being always formed at first at all temperatures down to −100° c., though it may freeze later. Both these surfaces soon "age" and then they act more like other common surfaces, dew being formed above about −30° c. to −40° c. and hoar frost at lower temperatures. A cadmium surface is similar to zinc, while a

polished aluminium surface is more like mica. All polished metal surfaces have a greater tendency to have dew deposited on them than crystalline surfaces of the same metal. When measuring saturation temperatures at which either dew or hoar frost may be formed and are difficult to distinguish, we find it best to cool the thimble sufficiently to freeze any deposit of dew and then to allow it to warm up until this ice deposit has nearly, but not quite, evaporated.

Various experiments were made to try to "seed" the thimble surface so that hoar frost would be formed instead of dew, but without success. Crystalline quartz dust appeared to be a likely substance since both quartz and ice form hexagonal crystals. It was, however, quite without effect, and even if quartz crystals were dropped into supercooled water, they had no more tendency to cause crystallization than glass dust. It has been suggested that quartz dust in the atmosphere would form natural nuclei on which water vapour would sublime at low temperatures to form ice crystals, but apparently this is not the case.

It is of some interest to note that dew can be formed on most surfaces at any temperature above about $-100°$ C. if the proper precautions are taken—the dew, of course, often freezes rapidly after being formed. Thus, if the thimble of a dew-point hygrometer be cooled to a low temperature in a current of air which is so dry that no deposit is formed and then a puff of damp air is admitted, dew will be formed on the thimble. Even if the thimble has a number of ice crystals scattered over it dew will be formed on the surface between the crystals.

A further point of interest is that two supercooled water droplets may grow on the thimble until they touch and coalesce and they will still remain liquid. Thus it appears probable that supercooled cloud and rain droplets which may collide in the air will not necessarily freeze as has sometimes been suggested.

(*b*) When the thimble of a dew-point hygrometer is cooled and hoar frost is deposited, it is almost always found that the hoar frost is not deposited immediately the thimble temperature falls below the frost-point, but that it has to be cooled nearly to the dew-point before any deposit is formed. The temperature of the thimble when the first trace of frost appears is therefore not a good indication of the frost-point even if the thimble be cooled very slowly. Nor is the mean of the temperatures when a deposit is first seen on cooling and last seen on warming an accurate measure of either dew- or frost-point. One must find the temperature of the thimble when the deposit is neither growing nor evaporating.

(*c*) At the lower temperatures, say $-70°$ C. and $-80°$ C., the amount of water vapour in the air is so small and the vapour pressure is so low, that both the growth and evaporation of the deposit are very slow. If an eye-observation instrument is used, it is necessary to wait at least a minute with the thimble held at a constant temperature to be sure whether the deposit is just growing or just evaporating. This makes the measurement very tedious and the photo-electric depo it indicator has the great advantage that a small change in the amount of the deposit can be quickly seen.

(*d*) As drier and drier air is used in a frost-point hygrometer the nature of the deposit gradually changes. For frost-points above about $-30°$ C. the deposit, if ice, consists of relatively large white crystals which can easily be seen with a magnifying lens. At lower temperatures the deposit consists of a uniform layer of fine crystals, and below a temperature of about $-70°$ C. these crystals are so fine that a thin deposit has a definite blue colour, while at about $-80°$ C.

the deposit is often difficult to see and appears as little more than a faint bloom on the thimble. At still lower temperatures no visible deposit is formed at all, and the water vapour is apparently condensed as a uniform glassy layer. Even if the thimble be first coated with a thin layer of hoar frost and then cooled to say $-120°$ C. in air of frost-point about $-100°$ C., the deposit does not appear to grow, even if it is measured by the very sensitive photoelectric deposit indicator described later. That a layer of glassy ice is actually formed in the above case when the thimble is cooled below the frost-point in very dry air is shown by the following experiment. The thimble is first cooled to, say, $-120°$ C. in a stream of extremely dry air (obtained by passing air through a spiral, cooled in liquid air), then air of frost-point about $-70°$ C. is passed for a minute or two, during which time there is no visible change in the thimble surface ; the extremely dry air is then again passed and the thimble slowly warmed. As the thimble warms up to a temperature of $-90°$ C. to $-80°$ C. a deposit appears to form although it is in air of frost-point far below this temperature, and if held at this temperature the deposit slowly evaporates again. There seems no doubt that an invisible deposit of glassy ice is deposited at the very low temperature and that as it warms up to about $-90°$ C. it begins to turn into crystalline ice. The formation of glassy ice at low temperatures was confirmed by an experiment made in a vacuum. A molecular beam of water vapour was directed against a cooled polished gold surface which was illuminated with monochromatic light. When the temperature of the surface was above about $-90°$ C. a visible deposit was formed as usual and no optical phenomena were observed. When the temperature of the surface was below $-90°$ C. Newton's rings were observed, and these moved as the layer of glassy ice increased in thickness. If, after deposition at very low temperatures, the surface was slowly warmed, the deposit changed on passing a temperature of about $-90°$ C. to a translucent layer and the rings disappeared.

This phenomenon of the formation of glassy ice is unfortunate since it appears to set a lower limit to the frost-point which can be measured with such an instrument. Though this limiting temperature is low, about $-90°$ C., frost-points have been observed in the stratosphere almost down to the limiting values, and we should certainly like to be able to measure still drier air.

§ 3. DESCRIPTION AND USE OF THE HYGROMETERS

The essential purpose of the instrument is to ventilate with the air under measurement a surface covered with dew or frost, and to find the temperature of the surface at which the dew or frost is in equilibrium with the air. The prime requirements of the instrument are therefore (*a*) a properly ventilated surface, the temperature of which can be easily and smoothly controlled and accurately measured, and (*b*) satisfactory arrangements to determine whether the deposit of dew or frost is changing. In the visual hygrometer any change in the amount of the deposit is judged visually, and the eye is given as much assistance as possible.

The test surface is formed by the top surface of the thimble, seen in the diagram (figure 3). It is a relatively massive piece of pure aluminium and is, for special reasons, anodized and dyed black. Its thermal inertia prevents too rapid changes in temperature. Its temperature is adjusted by balancing a controlled cooling against natural heating, which is provided by the heat of the air stream ventilating the top of the thimble and by thermal conduction from

other parts of the instrument. The conduction is kept at a low value by supporting the thimble on a sheet of paxolin $\frac{1}{8}$" thick, and surrounding the skirt with a sleeve which prevents heating of the skirt by convection. An electrical

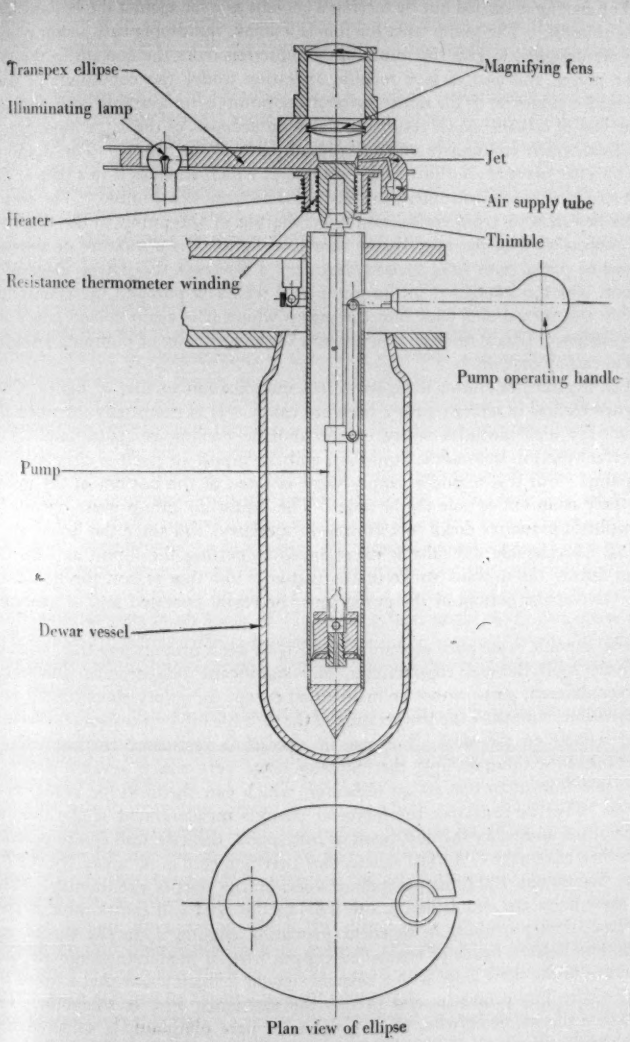


Figure 3. Diagram showing the principle of the eye-observation frost-point hygrometer. (*Reproduced, by kind permission, from the 'Proceedings of the Royal Society'.*)

heater is wound on this sleeve for use when natural heating is insufficient. Its use is discouraged by introducing a switch for the electric supply which must be held on during the period that additional heating is required.

The cooling is carried out by forcing a coolant as a jet against the hollow base of the thimble. The pump used for this is a small, hand-operated piston pump carefully designed to give the minimum conduction from the coolant to the top of the pump, and also to give reliable operation under the conditions of use. Excessive conduction to the upper parts of the pump is undesirable both because of the loss of coolant which results and also on account of the large amounts of hoar frost which accumulate on the exposed parts of the pump. The pump is held by a flat bayonet joint in a flange at the top, which attaches it to a part of the main structure of the instrument entirely separate from the thimble. The object of this design is to avoid conduction from the top of the pump to the thimble. The coolant used in the hygrometer shown in figure 3 is a mixture of alcohol, acetone or petrol with solid carbon dioxide. This cools the petrol, alcohol or acetone, which is contained in the pump and which is pumped up against the thimble and then drains back into the pump where it is again cooled ready for recirculation. This is necessary on account of the difficulty of pumping a sludge of solid carbon dioxide.

The hygrometer shown in figure 4 is a modification of that of figure 3, redesigned for use in aircraft with a pressure cabin. It is essentially intended for use at very high altitudes where carbon dioxide cooling is inadequate. The refrigerant used in this case is liquid air, and the liquid air itself is circulated by the pump. For this reason a supply valve is fitted at the bottom of the pump and there is an outlet hole in the side. The liquid air pump must operate at atmospheric pressures down to 150 mm. of mercury, and since the liquid air is boiling, considerable difficulty is experienced in getting the liquid air into the pump during the upward stroke of the piston. For this reason the liquid air inlet valve at the bottom of the pump must be easily operated and of generous size.

The thimble is of pure aluminium and is of such proportions that, with its relatively high thermal conductivity, any significant temperature differences between different parts cannot be maintained except for a very short time. It is thus possible to measure the temperature of the thimble top by a resistance thermometer wound on the skirt. For use in aircraft, a resistance thermometer is preferable to a thermocouple thermometer since very much greater power is obtainable from it to operate an indicator, which can therefore be made more robust. Also the resistance thermometer permits measurement of the thimble temperature, and hence the dew-point or frost-point, directly, and this in practice is another advantage. A very satisfactory thermometer is obtained by cutting a very fine thread, 100 turns per inch or finer, on the skirt of the thimble. Two fine skew holes are also drilled in the walls of the skirt and the thimble is then anodized, thus providing an excellent insulating coating upon the thread and in the fine holes. Leads of enamelled copper wire of moderate gauge are then cemented in the skew holes with a bakelite cement in such a way that a tip of the copper wire just protrudes and to this the resistance wire is soldered. The resistance thermometer wire, of 50 μ diameter bare platinum, is soldered first to the projecting tip of one copper lead and then wound down the thread on the skirt to the other lead, to which it is soldered. The wire lies at the bottom

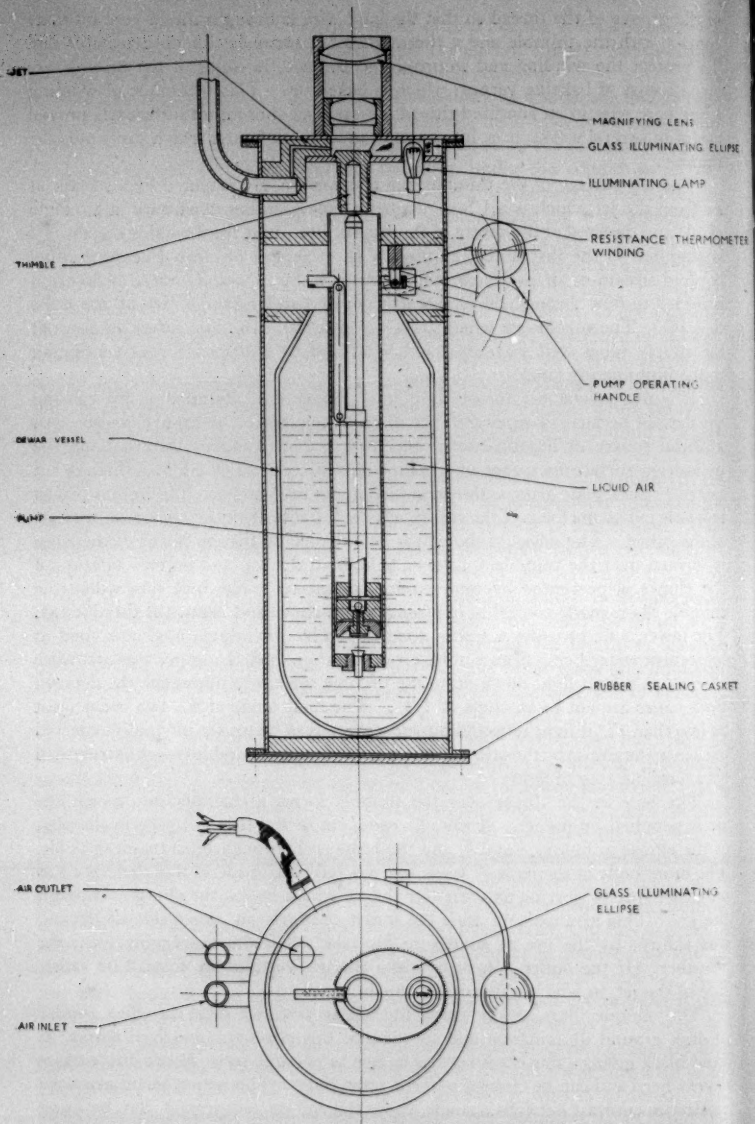
of the groove of the thread so that the bare wire is brought into a very intimate contact with the thimble and a thermometer is formed with no detectable lag. To protect the winding and improve the contact the sides of the thimble are given a coat of bakelite varnish which is baked on. This technique of winding a fine bare wire in an anodized thread cut upon an aluminium surface has proved useful in several ways; it is, for example, used in the heater which surrounds the thimble.

The top surface of the thimble, the test surface, is ventilated by a stream of air from the jet which is a 1 mm. diameter hole pointing downward at an angle of 10° and directed at the centre of the test surface. Air from outside the aircraft is supplied to the instrument through a short length of clean $\frac{1}{4}''$ copper tube. A rapid stream of air passes constantly through the tube and a small proportion is forced to flow through the hygrometer jet by reducing the outlet of the tube to 3/16". The surplus air which flows through the 3/16" hole serves to keep all the supply pipes well scavenged and is utilized to provide a forced ventilation to the illuminating lamp.

It is not satisfactory to use daylight as a source of illumination for viewing the deposit because in an aircraft the illumination is liable to rapid changes. An artificial source of illumination is therefore provided and is thrown upon the deposition surface by means of an elliptical reflector. The ellipse, which is cut out of $\frac{1}{4}''$ thick plate glass, is shown in plan in figures 3 and 4. The light is placed in a hole cut at one focus of the ellipse and the thimble fits in a conical seat in a hole at the other. The edges of the ellipse are silvered so that an image of the lamp is formed over the thimble. Loss of light from the top and bottom surfaces of the ellipse is prevented by total internal reflection. The hole into which the thimble fits is made conical in opposing senses above and below the thimble top. The lower cone provides a seating for the thimble top which must be a good fit to prevent water being drawn in by capillary action, and the upper cone provides a prism to refract light down on to the thimble surface to illuminate the deposit. Both cones are cut to an angle of $17\frac{1}{2}°$. The total angle of the two cones must be less than 37° if light reflected by one surface is to be totally internally reflected back into the ellipse by the other. If this is not done the thimble is seen surrounded by a dazzling ring of light.

The hole in the ellipse over the thimble forms a chamber into which the air is supplied by the jet. A slot, $\frac{1}{4}''$ wide, cut in the ellipse is seen in the plan of the ellipse in figures 3 and 4, and the brass jet is brought in through this slot. The brass body of the jet is $\frac{1}{4}''$ high, but it is only 0·2" wide so that there is a gap on either side to provide an outlet for the air which enters the chamber through the jet. This arrangement with the outlet channels on either side of the jet, was shown, by the use of smoke, to give the maximum scavenging from the chamber. If the outlet hole is opposite the jet, vortices are formed on either side of the jet, in which stale air can remain.

The oblique illumination which the ellipse provides gives an effect similar to dark ground illumination and to enhance this effect the anodized surface is dyed black giving a fine black surface as free as possible from sheen; this surface is very hard and can be cleaned without scratching. This arrangement gives the maximum contrast between a single ice crystal and its background, but, by using a small jet which fully ventilates only a narrow streak across the thimble the observed contrast is further increased, since the eye can compare the area of the



Plan view—cover plate removed.

Figure 4. Frost-point hygrometer for pressure cabin aircraft, diagrammatic illustration.
(Reproduced, by kind permission, from the ' Journal of the Royal Aeronautical Society '.)

jet in which deposit is at a maximum with the rest of the thimble which usually has little or no deposit. At very low temperatures where the individual ice crystals are too small to be seen, the eye is dependent upon this latter effect to perceive the deposit.

To assist the eye further a moderate amount of optical magnification is provided. It is necessary to see the whole of the thimble top at once in order to permit comparison between the deposit in the jet and the rest of the thimble. The The thimble top is 15 mm. diameter and a magnification of more than $\times 10$ would prevent simultaneous vision of the whole thimble. Actually a magnification of $\times 7$ is provided. It is not possible to use an ordinary magnifying glass or eyepiece for this purpose partly because of the absence of any particular eye point and partly because the bottom surface of the lower lens must form the top of the observation chamber, and it must therefore be as low as possible. It must be, however, lifted sufficiently for it not to be illuminated by the ellipse. If it is illuminated any dust or dirt on the bottom surface is very visible. The optical system adopted, figure 4, is therefore corrected after the manner of a Petzval portrait lens rather than a conventional eyepiece. The upper lens, an achromat of B.S.C. and E.D.F. glasses, is corrected for spherical aberration, it has slight positive coma and an inevitably large amount of positive astigmatism which would be very undesirable if uncorrected. This particular combination of glasses is used to permit achromatization with relatively long radii and therefore to permit a large diameter on the lens. This large size of the upper lens permits the eye to be moved at least 2″ from the upper lens and still retain a full view of the thimble. The lower lens is made an approximate achromat of H.C. and D.F. glasses, largely because the use of a doublet in this position greatly assists in the correction of the astigmatism of the upper lens. The combined system has no visible spherical aberration or coma. It has an almost flat tangential field for any reasonable position of the eye, and no visible chromatic aberration.

The combination of dark ground illumination and good quality optical magnification shows a crystal of ice such as is obtained at temperatures down to $-60°$ c. in brilliant contrast. At very low temperatures when the individual crystals cannot usually be seen the system shows up a deposit consisting of about 0·01 microgram of water as a bluish bloom in the line of the jet.

Misting of the lenses is apt to be troublesome since very cold air is circulating in the chamber under the lower lens while the cabin temperatures may be quite high. This is particularly troublesome in the pressurized instrument which is commonly used in a pressure cabin, probably with a cabin temperature of 15 or 20° c. and a high relative humidity, maintained by perspiration and exhalation by the aircraft crew, while the air under the lower lens may be $-80°$ c. (In the stratosphere it is possible with the same instrument to measure the CO_2 point, in which case the air in the test-chamber may be still colder.) This trouble is eliminated by making the tube which carries the upper lens of plastic so that thermal conduction is reduced and the upper lens is not cooled sufficiently for condensation to take place upon it. The lower lens mount, on the other hand, is deliberately made of brass or other good thermal conductor so that the low temperature of the test-chamber is readily conducted through the mount to the space between the lenses. Any water vapour in the space between the lenses then condenses upon the cold mount rather than upon the glass surface, which is slightly warmer due to the reduced conduction in the glass. To reduce con-

duction through the glass, the lower lens is made very thick. At a total thickness of the lower lens of 0·25″ some trouble was experienced in a pressure cabin, but by increase of thickness to 0·5″ condensation on the lenses was entirely eliminated.

Absorption of water by various materials is always a source of error as it is given up again when working with very dry air. Cleaned copper tubing is the most satisfactory material for supply pipe lines and flexible metal bellows must be used to take up any movement; rubber tubing must always be avoided. For the same reason it has been necessary to make the illuminating ellipses of glass, though the difficulties of manufacture are increased. Perspex is very suitable from the point of view of manufacture but it is useless if frost-points below $-30°$ c. are to be measured. Transpex II (polystyrene) is less satisfactory for manufacture, but can be used down to about $-60°$ c. The glass ellipse is satisfactory for use at all temperatures.

To use the instrument and obtain an accurate reading at low temperatures in a reasonable time considerable experience is required. The usual method is to pump slowly so that the thimble temperature falls at a reasonable rate, usually less than $\frac{1}{2}°$ c./sec., watching carefully for the formation of a deposit. The deposit does not necessarily begin to form immediately the frost-point is passed, so that the deposit may not be seen until after a quite appreciable overshoot. Immediately a deposit is seen, pumping is stopped and the thimble allowed to warm up one or two degrees, depending upon the rate of cooling which has been used, upon the actual frost-point, and upon the particular thimble, which may be very clean or may have been sensitized by past use. The thimble temperature is then held constant by careful pumping and the deposit examined intermittently to see if it is increasing or decreasing. With experience it is then easy to find two temperatures at which the deposit is (*a*) very slowly increasing, and (*b*) very slowly decreasing, at approximately the same rates, and the dew- or frost-point is taken as the mean. On the other hand, if care is not taken the deposit may be lost altogether or may become too heavy. In either case much time is lost in obtaining a suitable deposit again. In order to get accurate measurements the deposit should be very thin since if it becomes at all heavy, changes are difficult to observe. It is very important that a deposit should be maintained throughout the observation and changes in the deposit should be observed while the thimble temperature is held steady. The frost-point is *not* the mean of the temperatures at which the deposit comes and goes from a thimble whose temperature is fluctuating rapidly, and this tempting method of use must be strictly avoided.

Laboratory tests of the accuracy of the instrument. At frost-points down to $-50°$ c. the accuracy which is obtainable is largely determined by the accuracy with which the temperature of the thimble can be controlled, and it is not difficult to measure a frost-point of $-50°$ c. repeatedly to $-\frac{1}{2}°$ c. in the laboratory, since the temperatures of deposit increase and decrease may differ by less than, say, 2° c. This accuracy can be obtained in an aircraft if care can be taken, though usually it is convenient to accept a less accurate observation which can be made more quickly and with much less effort. The humidity of the air through which an aeroplane is flying sometimes varies quickly so that great accuracy is not necessary.

Tests were made of the fundamental accuracy of the hygrometer at $-78·5°$ c.

to determine (*a*) whether the frost-point as measured was the true saturation temperature, (*b*) whether there was any effect on the measured frost-point when the thimble surface was changed from gold to the black anodised aluminium surface, and (*c*) the probable error of an observation at this temperature, and whether this was different for the two kinds of deposition surface or whether there was any major difference between various experienced observers.

Air with a frost-point of −78·5° C. was produced by first passing it through large drying tubes of silica gel at a pressure of 120 lb./sq.in. and expanding to atmospheric pressure; this gives air with a frost-point of about −65° C. The air was then passed through two spirals of clean copper tube immersed in a bath of solid carbon dioxide and acetone, which, provided it is boiling at a reasonable rate, gives a satisfactory constant temperature of −78·5° C. The air from the spirals was next passed through a large glass-wool filter also immersed in a carbon dioxide and acetone bath; the filter is used to remove particles of snow which are carried along by the air-flow and its use is essential. Provided that the system is not expected to remove too much water and that there is no excessive pressure difference across the filter, this arrangement gives air of a very constant water content and its frost-point should be −78·5° C., the temperature of the baths. The air was finally passed to a hygrometer where its frost-point was measured. In the laboratory a copper-constantin thermocouple inserted into the thimble under the test surface was used to measure the thimble temperature, and for this experiment the other junction was placed in the cooling bath surrounding the filter so that the measured frost-point should be at zero E.M.F.

The results obtained in one set of readings are shown in table 2.

Table 2

Observer	Temperature difference, Thimble minus Bath: Deposit just growing (Galvanometer deflection)	Deposit just decreasing (Galvanometer deflection)	Mean
	Black aluminium thimble		
A.W.B.	−2·5 div.	+3·5 div.	+0·5 div.
B.C.	−1·5	+4·5	+1·5
G.M.B.D.	−0·5	+1·5	+0·5
		Mean	+0·8 div.
	Gold topped thimble		
A.W.B.	−2·5 div.	+1·5 div.	−0·5 div.
B.C.	−1·5	+1·5	0
G.M.B.D.	−1·5	+1·5	0
		Mean	−0·2 div.
	Black aluminium thimble		
G.M.B.D.	−0·5 div.	+3·5 div.	+1·5 div.
B.C.	−1·5	+1·5	0
		Mean	+0·7 div.

(*Note.* Thimble temperatures were held constant for about a minute at various whole divisions on the galvanometer, and at the end of the minute the observer decided whether the deposit had grown or decreased. The galvanometer had a zero error of 0·5 divisions, therefore all values are given to this accuracy. 1 div.∼1° C. Each complete observation took about five minutes.)

It will be seen that the frost-points as observed by the three observers agree to within about 1° c. and that there is no significant difference between a gold and a black anodized aluminium thimble, nor is there significant difference between the expected and measured frost-point. In general most observers have a slight preference for the black anodized thimble, and this is used for the final instrument mainly because the provision of a resistance thermometer is so much easier.

Air with a frost-point as low as −78·5° c. is usually only encountered at heights in the atmosphere where, owing to shortage of oxygen and to discomfort, it is impossible to make the observations with the care possible in the laboratory; the more accurate type of instrument described below should therefore be used at very low temperatures.

§4. FROST-POINT HYGROMETER WITH PHOTOELECTRIC DEPOSIT INDICATOR

The eye-observation frost-point hygrometer described in the previous section makes a very simple instrument and is easy to use when measuring air whose frost-point is above about −50° c. When it has to be used to measure the humidity of still drier air it suffers from serious disadvantages. Although it is possible to measure the frost-point of very dry air down to −80° c., the rate of deposition and evaporation of the hoar frost become very slow so that it is necessary to keep the thimble at a constant temperature for about a minute before it is possible to decide whether the deposit is increasing or decreasing. At −80° c. the deposit is very faint indeed. This makes the measurements very long and fatiguing to the observer and the accuracy is rather poor. An instrument with a photoelectric deposit indicator has been developed to remove these disadvantages, and though it is necessarily more complicated than the simple eye-observation instrument, it is so much more accurate, so much quicker, and so much less tiring to use, that its use is always worth while at very great heights. Also, since much less judgment is required, the observations may be made by a less highly trained observer. Photoelectric dew-point hygrometers have been described by other workers, including Moss (1934), Awbery and Griffiths (1933), Thornthwaite and Owen (1940), and others in the U.S. and Canada. The optical arrangements employed are generally unsuitable for the lowest frost points.

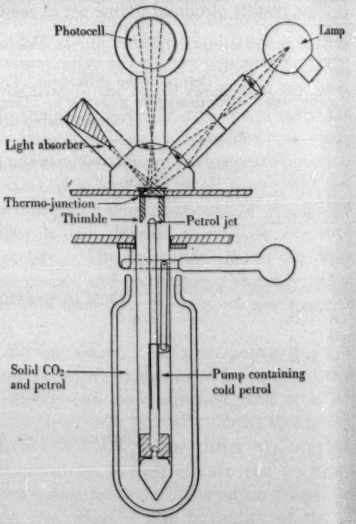


Figure 5. Diagram showing the principle of the frost-point hygrometer with photoelectric deposit indicator
(*Reproduced, by kind permission, from the 'Proceedings of the Royal Society'.*)

The principle of the instrument is seen from the schematic diagram

in figure 5. Light from a 36-watt lamp is concentrated by lenses on to the thimble. The thimble is made of copper with a gold plate soldered to its top, the gold surface being well polished or of aluminium with an anodized black top. The light is reflected from the thimble into a black absorbing box, through a slot slightly larger than the image of the lamp filament. A slot between the two lenses is also larger than the image of the filament and prevents the passage of all light except that directly from the filament. A diaphragm of suitable aperture cuts out reflections from the inside of the tube in which the lenses are mounted. As a result the inside of the chamber is fairly dark, except that an area about 8 mm. in diameter on the top of the thimble is strongly illuminated. A lens immediately above the thimble forms an image of this illuminated area at a hole through which the light passes to the photoelectric cell. The lenses are all ventilated by a stream of air from within the test-chamber to prevent any condensation on them, as they will tend to be warmer than the outside air.

If the surface of the thimble is clean and properly polished very little light is scattered from it on to the photocell. It is important both in the final polishing and in subsequent cleaning that all rubbing of the surface be only in the plane of the light since any minute scratches at right angles to this direction show up brightly. A small deposit of dew or hoar frost on the thimble causes a great

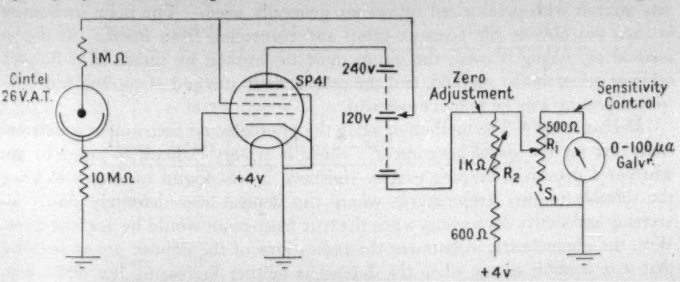


Figure 6.

relative increase in the amount of light reaching the photocell. The angle of the illuminating beam and the position of the photocell were fixed after careful trials using very faint deposits formed in very dry air. Somewhat different positions might be better for heavy deposits which appear white, but there is always ample light in these cases and the positions were chosen which were best for the very faint bluish deposits (see § 2).

The electric circuit is shown in figure 6. With a grid leak of resistance greater than 10^7 ohms the instrument is too sensitive, even when measuring air having a frost-point of $-80°$ c. The instrument is made to be of maximum sensitivity when used with frost-points of $-80°$ c. and when measuring higher frost-points the sensitivity is decreased by adjusting the resistance R_1. The indicating microammeter is never shorted by a resistance less than several times its own resistance so that it shall not be too sluggish. For full sensitivity the switch S_1 is opened. By adjusting the resistance R_2 the current through the indicating microammeter is made approximately zero when there is no deposit on the thimble. When using very dry air it is found that the change in the amount of

scattered light produced by a given small increase in the deposit is greatest for a moderately thin, but not very thin, deposit. With the photocell and amplification used, a deposit giving a reading of a few hundred microamperes is most sensitive. It is convenient to have an indicating galvanometer with high sensitivity in the middle of its range (when the current is about a hundred microamperes) and becoming less sensitive at each end of the scale.

The air to be measured is led in through a slot 10 mm. × 1 mm., the centre of the jet being directed at the centre of the thimble and making an angle of approximately 10° with the surface, as in the visual hygrometer. The air jet more than covers the illuminated area and is directed at right angles to the plane of the illumination.

The thimble may be cooled by pumping a jet of liquid air against the underside, as in the case of the visual instrument described above; or it may be cooled by being connected by a short, thin bar to a massive copper bar passing down into a Dewar flask of liquid air. In the latter case a heating coil is wound on the side of the thimble in a manner similar to the winding for the resistance thermometer already described. The current through the heating coil is controlled in order to keep the thimble at the required temperature. The photoelectric instrument is chiefly used when climbs are made to very great heights, in which case aircraft with pressurized cabins are generally used. The main instrument in then put outside the pressure-cabin and controlled from inside. If the jet method of cooling is used, the pump must be worked by means of a Bowden cable or other similar method, and the conductivity method of cooling with controlled heating may be more convenient.

Method of use.—The method of using the photoelectric instrument is different from that for the visual instrument. Since it is very difficult to judge by eye whether a deposit is keeping exactly constant, it was found necessary to bring the thimble to two temperatures where the deposit was definitely slowly increasing and slowly decreasing, when the true frost-point would be near the mean. With the photoelectric instrument the indications of the deposit are so sensitive that it is possible to say when the deposit is neither increasing nor decreasing, and the temperature of the thimble is adjusted to such a value and the temperature read at this condition. As the actual amount of the deposit is not important (provided it is not too heavy), many readings can be taken quickly one after the other with slightly different amounts of deposit. At least two independent measurements can be taken per minute so that a very reliable average can quickly be obtained.

Since the thimble cannot be seen by the observer there may be some doubt whether the deposit consists of dew or frost at temperatures between 0° and about −30° C. Since dew consists of very numerous fine droplets, while frost at these temperatures consists of much larger ice particles, a deposit of dew increases and decreases much more rapidly than one consisting of frost, and a skilled observer may judge the nature of the deposit from this. However, it is generally safer to cool the thimble sufficiently to be sure that the deposit has frozen, then to warm it until nearly all the ice has evaporated, leaving only a very thin deposit, and to work with this.

Measurements of frost-point can be made with this instrument down to the temperature at which the nature of the deposit changes from a crystalline to an invisible glassy layer.

The following critical tests of the instrument have been made in the laboratory with satisfactory results: —

A. To see what effect the thickness of the deposit had on the observed frost-point.

Observations were made using a very thin, a moderate and a very thick deposit, the order of the observations being shown by the figures within brackets (table 3).

Table 3

Observed frost-points (°c.)					
Thin deposit		Moderate deposit		Thick deposit	
			−57·2		−58·6
		(1)	−57·1	(2)	−57·8 ?
			−57·3		−58·6
		Mean	−57·2	Mean	−58·4
	−57·5		−58·2		−57·8
(3)	−57·5	(4)	−58·0	(5)	−58·0
	−57·3		−58·0		−58·2
Mean	−57·5	Mean	−58·1	Mean	−58·0
	−58·0		−58·3		−58·5
(6)	−58·1	(7)	−58·4	(8)	−58·3
	−58·3		−58·7		−58·7
Mean	−58·1	Mean	−58·5	Mean	−58·5
	−58·2				
(9)	−58·2		—		—
	−58·2				
Mean	−58·2				
Means	−57·9		−57·9		−58·3

The air supply was obtained by passing air over silica gel. It will be noticed that the air supply becomes slowly drier as the connecting pipes dry out. All the individual measurements are given so that the casual variations can be seen.

Since the thick and thin deposits used on the above test were outside the range of thickness usually used, no appreciable change of observed frost-point with deposit thickness is likely.

B. To see if the brightness of the illuminating lamp caused any change in the observed frost-point (e.g. through heating of the deposit on the thimble).

For this test the lamp was run alternately on 4 and 6 volts (the normal rating), while many measurements of the frost-point were made.

The means of the observed values were −50·7° c. for the 6-volt run and 50·2° c. for the 4-volt. Since the energy radiated by the lamp is much greater when on 6 volts than on 4 volts (which was the minimum on which the instrument would work satisfactorily), it may be taken that there is no appreciable effect.

C. To see if the rate of air-flow through the jet makes any difference to the observed frost-point.

The pressure on the jet was adjusted alternately to 4 in. of water and 24 in. with the following results (table 4).

Table 4

Observed frost-points (°c.)			
Pressure on jet			
24 in.		4 in.	
	−54·8		−55·3
(1)	−54·5	(2)	−55·2
	−54·8		−55·2
Mean	−54·7	Mean	−55·2
	−56·2		−56·2
(3)	−56·0	(4)	−56·0
	−56·0		−55·9
Mean	−56·1	Mean	−56·0
	−56·2		
(5)	−56·0		—
	−56·0		
Mean	−56·1		
Means	−55·6		−55·6

Note that the air supply is becoming drier during the first part of the test, probably due to the tubing drying out.

ACKNOWLEDGMENTS

This work was done while two of us (A.W.B. and B.C.) were on the staff of the Meteorological Office, and we are indebted to the Director both for permission to publish this paper and for his personal interest in the progress of the work.

Figures 1, 2, 3 and 5 are reproduced by permission of the Council from the *Proceedings of the Royal Society*, and figure 4 from the *Journal of the Royal Aeronautical Society*, by permission of the Council of that Society.

REFERENCES

Awbery and Griffiths, 1933, *The Measurement of Humidity in closed spaces*. D.S.I.R., Food Investigation Report No. 8 (London : H.M.S.O.).
Diamond, Hinman, Dunmore and Lapham, 1940, *Bur. Stand. J. Res., Wash.*, **25**, 327.
Dymond, 1947, *Proc. Phys. Soc.*, **59**, 645.
Ebert, 1937, *Z. f. Kälte-Ind.*, **44**, 127.
Glückauf, 1945, *Quart. J. R. Met. Soc.*, **71**, 110.
Gregory, 1947, *Instrum. Practice*, **1**, 367.
Griffiths and Awbery, 1935, *Proc. Phys. Soc.*, **47**, 684.
Harrison and Brewer, 1944, *Meteorological Research Committee Reports*, M.R.P. 205.
Moss, 1934, *Proc. Phys. Soc.*, **46**, 450.
Simons, 1936, *Proc. Phys. Soc.*, **48**, 136.
Thornthwaite and Owen, 1940, *Mon. Weath. Rev., Wash.*, **68**, 315.

The Cavity Resonator Method of Measuring the Dielectric Constants of Polar Liquids in the Centimetre Band

By C. H. COLLIE, J. B. HASTED and D. M. RITSON
Clarendon Laboratory, Oxford

MS. received 24 *January* 1947

ABSTRACT. A method is described of determining the refractive index n of a polar liquid in the centimetre band by measuring the resonance curve of an H_{01} resonator with an axial capillary of liquid. An independent value of ϵ is required. The method has been tried out on water at 21° c. Three independent measurements gave $n = 6{\cdot}25$, $6{\cdot}30$, $6{\cdot}35$, so that the method is capable of $\pm 1\%$ accuracy.

§ 1. INTRODUCTION

The most convenient way of measuring the dielectric properties of nearly ideal dielectrics (power factor 10^{-3}) is the cavity resonator method introduced by Willis Jackson (1941), using an E_{010} resonator, and Penrose (1946), using an H_{01} resonator. The latter method is in some respects more convenient at the highest frequencies ($\lambda = 1$ cm.), since it is possible to work with an oscillator of fixed frequency.

It is our purpose to examine whether these methods can be used when the dielectric is a polar liquid such as water, which is far from being a perfect dielectric (power factor $\frac{1}{2}$), and furthermore, which has to be maintained in the appropriate form by a containing vessel. If the problem is considered by an approximate method due to Bethe (1942), it is easily seen that it is not possible to arrange the specimen so as to obtain a large frequency shift and still maintain a sharp enough resonance curve for the frequency shift to be measured.

Two main difficulties are encountered in a straightforward application of the resonator method. In the case of the E_{010} resonator, in which the electric field is greatest on the axis, the loss is such that at 1 cm. the specimen must be enclosed in a capillary of less than 0·05 mm. diameter, which is impracticable. Moreover, it is difficult to achieve 1% accuracy with a fixed resonator and tunable oscillator. In the case of the H_{01} resonator, the difficulty is one of introducing the liquid specimen in the form of a flat disc of thickness 0·2 mm. These difficulties can to a great extent be overcome by using an H_{01} resonator with an axial capillary filled with liquid. The electric field is zero on the axis of the resonator so that a capillary tube large enough for accurate measurement can be used without damping the resonator so much that the shape of the resonance curve cannot be determined. The cylindrical symmetry makes it possible to solve Maxwell's equations exactly, so that the small corrections, which have to be made to take into account the effect of the glass wall of the capillary tube, can be evaluated as accurately as one wishes.

§ 2. THE RESONANCE CURVE AND THE PROPAGATION CONSTANT

The shape of the response curve of a resonator can be obtained by considering the course of a wave travelling back and forth in the resonator (e.g. Collie (1944)).

Consider the H_{01} resonator of length $l/2$ shown in figure 1.

If precautions are taken that new types of wave are not created and subsequently detected, the effect of the plunger and the blank face and detector at the input end can be represented by complex reflexion coefficients R_p and R_i respectively.

Then the electric field at the detector due to waves travelling from left to right is

$$A(1+R_pR_ie^{-\gamma l}+(R_pR_i)^2e^{-2\gamma l}+\ldots.)=\frac{A}{1-R_pR_ie^{-\gamma l}},$$

while the field at the detector of waves travelling from right to left is

$$Be^{-\gamma l}R_p(1+R_pR_ie^{-\gamma l}+\ldots.),$$

so that the total electric field E at the detector is given by

$$E=\frac{A+BR_pe^{-\gamma l}}{1-R_pR_ie^{-\gamma l}},$$

in which γ is the propagation constant $\alpha+j\beta$ of the wave in the guide.

Near resonance let

$$l=l_0+\Delta l,$$

$$\beta l_0=2n\pi-\phi,$$

$$R_pR_i=e^{-r-j\phi},$$

$$BR_pe^{-\gamma l_0}=C.$$

Then

$$E=\frac{A+Ce^{-\gamma\Delta l}}{1-e^{-r-j\phi-\gamma l}},$$

$$=\frac{A+Ce^{-\gamma\Delta l}}{(r+\alpha l)-j\beta\Delta l},$$

in which A can be treated as real without affecting the argument.

Then $|E|$ is given by

$$|E|=\frac{F+G\Delta l}{\{(r+\alpha l_0)^2+\beta^2\Delta l^2\}^{1/2}}.$$

This is a conventional resonance curve made slightly asymmetric by the small linear term $G\Delta l$.

The error involved in treating it as a conventional Q curve and measuring $\frac{1}{2}$(full width) when $E=E_{\max}/\sqrt{2}$ is quite small; if the degree of asymmetry is 5% at the half value width, the error in estimating $(r+\alpha l)$ is only 0·15%. A more satisfactory graphical treatment is to plot the observed value of Δl^2 against the square of the reciprocal of the mean value of the voltage response. The constant $r+\alpha l$ is readily obtained from the slope and intercept of the resulting straight line.

A typical response curve of a damped resonator is shown in figure 2 together with the linear plot described above. It can be seen that an excellent straight line is obtained: this not only justifies the elementary theory just outlined but provides a good check on the linearity of the receiving system.

The quantities required in theoretical calculations are α and β. These can be obtained if two successive resonances are measured. The amount the plunger has to be withdrawn is a direct measure of the wavelength λ_g in the guide and so of $\beta=2\pi/\lambda_g$. Since r is the same for both positions of the plunger, α can be obtained by subtraction from the measured values of $r+\alpha l_0$ and $r+\alpha(l_0+\lambda_g/2)$. Thus, provided the resonance curve is not too flat for accurate measurement, one can determine γ, the propagation constant in the guide, experimentally from resonance data.

§3. PROPAGATION IN A COMPOSITE GUIDE

It remains to obtain the relationship between the propagation constant γ in a guide partly filled with dielectric and the constants of the material. A general treatment of this problem is given by Pincherle (1944). Consider the propagation of an H_{01} wave down to the composite guide shown in figure 3, consisting of a meta

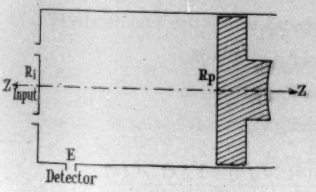


Figure 1.

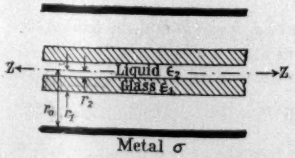


Figure 3.

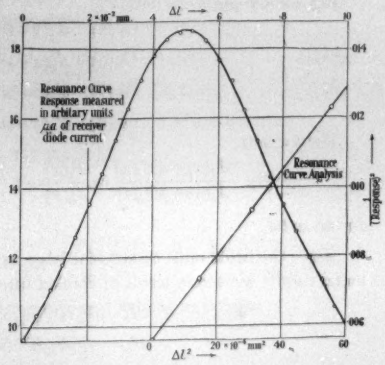


Figure 2. Resonance curve plotted and analysed as a straight line.

cylinder of internal radius r_0, with a glass capillary tube of external radius r_1, and internal radius r_2, lying along its axis. Let σ be the conductivity of the metal wall and ϵ_1 and ϵ_2 the complex dielectric constant of the glass and liquid respectively. Then the field equations for an H_{01} wave in cylindrical coordinates ρ, θ, z are *

$$H_z=k^2[aJ_0(k\rho)+bY_0(k\rho)], \qquad \ldots\ldots(1)$$

$$H_\rho=-\gamma k[aJ_1(k\rho)+bY_1(k\rho)], \qquad \ldots\ldots(2)$$

$$E_\theta=\frac{j\omega\mu k}{c}[aJ_1(k\rho)+bY_1(k\rho)], \qquad \ldots\ldots(3)$$

$$k^2-\gamma^2=\frac{\omega\epsilon\mu}{c^2}. \qquad \ldots\ldots(4)$$

The arbitrary complex coefficients a and b will have different values (cd, ef, gh) in each medium which must be chosen so that the tangential electric and magnetic fields are continuous at each interface.

* Lamont, *Wave Guides* (Methuen), p. 26; his $\beta=-j\gamma$.

One notes that γ must be the same for all four sections of the wave since these boundary conditions must continue to hold as the wave is propagated down the guide. Since γ is a measured quantity, the appropriate value of k (k_M, k_A, k_G, k_L) can be determined from equation (4) except in the inner region which is filled with the liquid whose dielectric constant ϵ_2 is being measured.

Approximations

The following well known approximations will be used:

For $a<0{\cdot}01$

$$\text{A(1)} \quad J_1(3{\cdot}831+a)\simeq aJ_1'(3{\cdot}831),$$
$$\text{A(2)} \quad J_0(3{\cdot}831+a)\simeq J_0(3{\cdot}831),$$
$$\text{A(3)} \quad Y_0(3{\cdot}831+a)\simeq Y_0(3{\cdot}831),$$
$$\text{A(4)} \quad Y_1(3{\cdot}831+a)\simeq Y_1(3{\cdot}831).$$

For $x\ll 3{\cdot}8$ and $a/x<0{\cdot}01$

$$\text{A(5)} \quad J_1(x+a)\simeq J_1(x),$$
$$\text{A(6)} \quad J_0(x+a)\simeq J_0(x),$$
$$\text{A(7)} \quad Y_1(x+a)\simeq Y_1(x),$$
$$\text{A(8)} \quad Y_0(x+a)\simeq Y_0(x).$$

For $b<0{\cdot}01$

$$\text{A(9)} \quad \frac{J_1(x)+bY_1(x)}{J_0(x)+bY_0(x)}\simeq\frac{J_1(x)}{J_0(x)}+b\frac{Y_1(x)\,.\,J_0(x)-Y_0(x)\,.\,J_1(x)}{[J_0(x)]^2}.$$

In the metal

Since the fields must be zero for large values of ρ, $b=-ja$ and the field in the metal can be written in terms of Hankel functions as

$$H_z=ak_M^2H_0^{(2)}(k_M\rho)e^{-\gamma z}, \qquad \ldots\ldots(5)$$
$$H_\rho=-a\gamma k_MH_1^{(2)}(k_M\rho)e^{-\gamma z}, \qquad \ldots\ldots(6)$$
$$E_\theta=\frac{aj\omega k_M\mu}{\mathbf{c}}H_1^{(2)}(k_M\rho)e^{-\gamma z}, \qquad \ldots\ldots(7)$$

in which k_M denotes k in the metal given by

$$k_M^2=\frac{4\pi\sigma\omega}{\mathbf{c}^2}+\gamma^2\simeq\frac{4\pi\sigma\omega}{\mathbf{c}^2}. \qquad \ldots\ldots(8)$$

Since the value of γ is never very different from that of an empty guide with perfectly conducting walls it is convenient to write our expressions in terms of their deviation from the propagation constant G of an ideal empty guide, given by

$$G^2=\frac{3{\cdot}831^2}{r_0^2}-\omega^2/\mathbf{c}^2.$$

The wave in the actual empty guide will be given by

$$H_z=a_0\,.\,{}_0k_A\,.\,J_0({}_0k_A\rho), \qquad \ldots\ldots(9)$$

in which the subscript zero indicates values for an empty guide.

From (8) it follows that $k_M\simeq{}_0k_M$, so that equating E_θ and H_z at the metal/air interface of an empty guide, one has

$$\frac{1}{k_Mr_0}\frac{H_1^{(2)}(k_Mr_0)}{H_0^{(2)}(k_Mr_0)}=\frac{1}{{}_0k_Ar_0}\,.\,\frac{J_1({}_0k_Ar_0)}{J_0({}_0k_Ar_0)}. \qquad \ldots\ldots(10)$$

Let $$_0k_Ar_0 = 3{\cdot}831 + \chi_1 + j\chi_2, \qquad \ldots\ldots(11)$$

in which the χs represent the effect of the finite conductivity of the wall and are very small.

Let $$\gamma_0 = G + \gamma_{01} + j\gamma_{02}, \qquad \ldots\ldots(12)$$

so that from (4), (11), (12) one has

$$\chi_1 + j\chi_2 \simeq \frac{r_0^2}{3{\cdot}831} G(\gamma_{01} + j\gamma_{02}), \qquad \ldots\ldots(12a)$$

and substituting in (10)

$$\frac{1}{k_Mr_0} \cdot \frac{H_1^{(2)}(k_Mr_0)}{H_0^{(2)}(k_Mr_0)} \simeq \frac{r_0^2}{(3{\cdot}831)^2} \cdot G \cdot \frac{J_1'(3{\cdot}831)}{J_0(3{\cdot}831)}(\gamma_{01} + j\gamma_{0}). \qquad \ldots\ldots(13)$$

The air metal boundary

The boundary conditions are expressed by the equations

$$ak_M^2H_0^{(2)}(k_Mr_0) = k_A^2[cJ_0(k_Ar_0)] + dY_0(k_Ar_0), \qquad \ldots\ldots(14)$$

$$ak_MH_1^{(2)}(k_Mr_0) = k_A[cJ_1(k_Ar_0)] + dY_1(k_Ar_0). \qquad \ldots\ldots(15)$$

These can be solved for d/c by division; k_Ar_0 is nearly 3·831, so writing

$$k_Ar_0 = 3{\cdot}831 + \phi_1 + j\phi_2$$

and using approximations A (1), A (2), A (3), A (4) and neglecting $(d/c) \cdot Y_0(3{\cdot}831)$ compared with $J_0(3{\cdot}831)$, one has

$$\frac{1}{k_Mr_0}\left(\frac{H_{(1)}^{(2)}(k_Mr_0)}{H_{(0)}^{(2)}(k_Mr_0)}\right) \simeq \frac{1}{3{\cdot}831}\left[(\phi_1 + j\phi_2) \cdot \frac{J'(3{\cdot}831)}{J_0(3{\cdot}831)} + \frac{d}{c} \cdot \frac{Y_1(3{\cdot}831)}{J_0(3{\cdot}831)}\right],$$

substituting for the L.H.S. from equation (13) and writing as before

$$\gamma = G + \gamma_1 + j\gamma_2,$$

so that $$G(\gamma_1 + j\gamma_2) \simeq \frac{3{\cdot}831}{r_0^2}(\phi_1 + j\phi_2)$$

as in (12 *a*), one finds that

$$\frac{d}{c} \simeq -\frac{J_1'(3{\cdot}831)}{Y_1(3{\cdot}831)} \cdot \frac{r_0^2G}{3{\cdot}831} \cdot (\gamma - \gamma_0) = u(\gamma - \gamma_0)$$

with $$u = \frac{-J_1'(3{\cdot}831)}{Y_1(3{\cdot}831)} \cdot \frac{r_0^2G}{3{\cdot}831}, \qquad \ldots\ldots(16)$$

u has no real part since G is purely imaginary.

The glass air surface

The boundary conditions are

$$k_A^2[cJ_0(k_Ar_1) + dY_0(k_Ar_1)] = k_G^2[eJ_0(k_Gr_1) + fY_0(k_Gr_1)], \qquad \ldots\ldots(17)$$

$$k_A[cJ_1(k_Ar_1) + dY_1(k_Ar_1)] = k_G[eJ_1(k_Gr_1) + fY_1(k_Gr_1)], \qquad \ldots\ldots(18)$$

so that $$\frac{1}{k_A} \cdot \frac{J_1(k_Ar_1) + (d/c)Y_1(k_Ar_1)}{J_0(k_1r_A) + (d/c)Y_0(k_Ar_1)} = \frac{1}{k_G}\frac{J_1(k_Gr_1) + (f/e)Y_1(k_Gr_1)}{J_0(k_Gr_1) + (f/e)Y_0(k_Gr_1)}, \qquad \ldots\ldots(19)$$

which can be solved for f/e since d/c can now be looked upon as known.

Making use of the relationship $J_1(x)Y_0(x)-J_0(x)Y_1(x)=2/\pi x$ one finds that

$$\frac{k_G}{k_A}\left[\frac{J_1(k_A r_1)}{J_0(k_A r_1)}-\frac{d}{c}\cdot\frac{2}{J_0^2(k_A r_1)\,.\,\pi k_A r_1}\right]\simeq\frac{J_1(k_A r_1)}{J_0(k_A r_1)}\,.\,-\frac{f}{e}\frac{2}{J_0^2(k_G r_1)\,.\,\pi k_G r_1}. \qquad \ldots\ldots(20)$$

This may be further reduced using A (5, 6, 7, 8), which are valid if the glass is nearly loss-free, so that $\epsilon'_G > \epsilon''_G/100$.

For the Corning 707 low-loss glass used, this approximation is justified.

One finds
$$\frac{f}{e}=v+w\,d/c \qquad \ldots\ldots(21)$$

with
$$v=-\pi\bar{k}_G r_1\left[\frac{\bar{k}_A}{\bar{k}_G}\cdot\frac{J_1(\bar{k}_A r_1)}{J_0(\bar{k}_A r_1)}-\frac{J_1(\bar{k}_G r_1)}{J_0(\bar{k}_G r_1)}\right]J_0(\bar{k}_G r_1), \qquad \ldots\ldots(22)$$

$$w=\frac{\bar{k}_G^2}{\bar{k}_A^2}\cdot\frac{J_0^2(\bar{k}_G r_1)}{J_0(\bar{k}_A r_1)} \qquad \ldots\ldots(23)$$

in which $\bar{k}_G$ and $\bar{k}_A$ denote the real parts of k_G and k_A respectively.

The glass liquid surface

The boundary conditions are

$$k_G^2[eJ_0(k_G r_2)+fY_0(k_G r_2)]=k_L^2 gJ_0(k_L r_2), \qquad \ldots\ldots(24)$$

$$k_G[eJ_1(k_G r_2)+fY_1(k_G r_2)]=k_L gJ_1(k_L r_2), \qquad \ldots\ldots(25)$$

there being no term $hY_0(k_L r)$ since the fields must be finite on the axis.

Using A (9) and neglecting the glass as before, one has

$$\frac{1}{k_L r_2}\cdot\frac{J_1(k_L r_2)}{J_0(k_L r_2)}\simeq\frac{1}{\bar{k}_G r_2}\left\{\frac{J_1(\bar{k}_G r_2)}{J_0(\bar{k}_G r_2)}+\frac{\frac{f}{e}[Y_1(\bar{k}_G r_2)J_0(\bar{k}_G r_2)-Y_0(\bar{k}_G r_2)J_1(\bar{k}_G r_2)}{J_0^2(\bar{k}_G r_2)}\right\}, \qquad \ldots\ldots(26)$$

$$\frac{1}{k_L r_2}\cdot\frac{J_1(k_L r_2)}{J_0(k_L r_2)}=\frac{1}{\bar{k}_G r_2}\cdot\frac{J_1(\bar{k}_G r_2)}{J_0(\bar{k}_G r_2)}+\frac{f}{e}\cdot\frac{2}{\pi(\bar{k}_G r_2)^2}\cdot\frac{1}{J_0^2(\bar{k}_G r_2)}, \qquad \ldots\ldots(27)$$

$$=y+\frac{f}{e}z,$$

so that
$$\frac{1}{k_L r_2}\frac{J_1(k_L r_2)}{J_0(k_L r_2)}\simeq y+zv+zwu(\gamma-\gamma_0). \qquad \ldots\ldots(28)$$

Determination of the refractive index n

From equation (4) we have

$$k_L^2-\gamma^2=\frac{\epsilon_L\omega^2}{c^2}=(n-j\kappa)^2\frac{\omega^2}{c^2}.$$

For polar liquids ($\epsilon_L \gg 1$) it is sufficiently accurate to write

$$k_L\simeq(n-j\kappa)\omega/c.$$

Substituting into equation (28) and equating real and imaginary parts

$$\mathfrak{Im}\left[\frac{c}{\omega r_2(n-j\kappa)}\cdot\frac{J_1(\omega r_2(n-j\kappa)/c)}{J_0(\omega r_2(n-j\kappa)/c)}\right]=zwu\,.\,\mathfrak{R}(\gamma-\gamma_0), \qquad \ldots\ldots(29)$$

$$=zwux,$$

$$\mathfrak{R}\left[\frac{c}{\omega r_2(n-j\kappa)}\cdot\frac{J_1(\omega r_2(n-j\kappa)/c)}{J_0(\omega r_2(n-j\kappa)/c)}\right]=y+zv+zwu\,\mathfrak{Im}(\gamma-\gamma_0),$$

$$\ldots\ldots(29\,a)$$

since u is a pure imaginary and v, ω, y, z are all real.

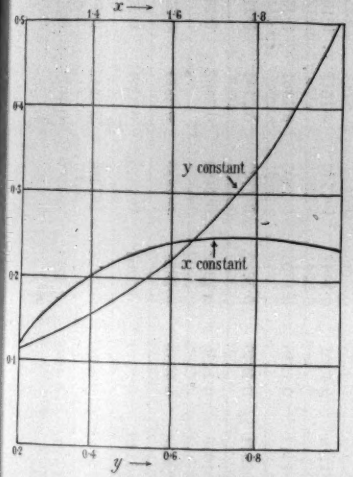


Figure 4. Values of $\mathfrak{Im}\left[\frac{1}{(x-iy)}\cdot\frac{J_1(x-iy)}{(J_0x-iy)}\right]$ for $x=1{\cdot}6$ and y varying and for $y=0{\cdot}5$ and x varying.

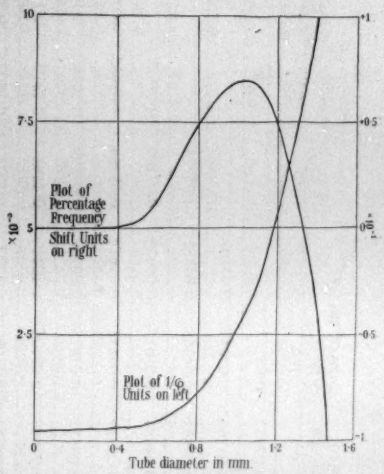


Figure 5. Approximate variation of $1/Q$ and percentage frequency shift with internal diameter of capillary, calculated and roughly checked experimentally for a typical polar liquid (water at 0° c., $n=5$, $k=2{\cdot}8$).

Since the value of κ *is* known from an independent experiment, equation (29 can be solved numerically for n using tables of

$$\mathfrak{Im}\frac{1}{z}\frac{J_1(z)}{J_0(z)}=\mathfrak{Im}(Q)$$

where $z=x-jy$.

If n were known independently, this could be used as a measurement of κ. The above function has been evaluated and is given in table 1, together with the real part of the function in table 2. The real part is included for information only, since it is not normally used, as the imaginary part of $Y-Y_0$, i.e. the frequency change of the resonator, cannot be determined experimentally with sufficient accuracy. This will be apparent from figure 5, where frequency shift and $1/Q$ are plotted against capillary tube radius, for a typical polar liquid. The error in calculating n from Y assuming that there is no frequency shift is in general

only a fraction of 1%, but originally a correction, computed from table 2, was included in the calculations. The behaviour of the frequency shift, which is in the normal direction for small tubes, and negative for large tubes, is interesting, and has been verified experimentally; it is probably due to the fact that the field

Table 1. Values of $\mathfrak{Im}(Q)$

$x \downarrow$ $y \rightarrow$	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
0·5	0	0·006753	0·01347	0·01970	0·02587	0·03112					
0·6	0	0·008421	0·01685	0·02464	0·03198	0·03847	0·03847				
0·7	0	0·01038	0·02052	0·03041	0·03895	0·04665	0·05381	0·05974			
0·8	0	0·01251	0·02466	0·03615	0·04684	0·05615	0·06433	0·07100	0·07629		
0·9	0	0·01503	0·02959	0·04317	0·05579	0·06651	0·07612	0·08364	0·08983	0·0942	
1·0	0	0·01803	0·03526	0·05163	0·06627	0·07875	0·08958	0·09799	0·1046	0·1091	0·1122
1·1	0	0·02153	0·04237	0·06151	0·07846	0·09304	0·1051	0·1144	0·1210	0·1257	0·1284
1·2	0	0·02596	0·05085	0·07364	0·09353	0·1100	0·1232	0·1331	0·1402	0·1443	0·1464
1·3	0	0·03135	0·06154	0·08874	0·1116	0·1305	0·1447	0·1550	0·1615	0·1650	0·1661
1·4	0	0·03881	0·07530	0·0176	0·1343	0·1550	0·1703	0·1805	0·1861	0·1885	0·1878
1·5	0	0·04838	0·09326	0·1321	0·1631	0·1840	0·2012	0·2102	0·2141	0·2142	0·2117
1·6	0	0·06159	0·1178	0·1646	0·1998	0·2236	0·2382	0·2450	0·2462	0·2431	0·2374
1·7	0	0·08068	0·1524	0·2090	0·2480	0·2716	0·2831	0·2855	0·2821	0·2745	0·2646
1·8	0	0·1095	0·2027	0·2711	0·3118	0·3319	0·3368	0·3318	0·3216	0·3079	0·2932
1·9		0·1558	0·2801	0·3592	0·3969	0·4069	0·3998	0·3841	0·3642	0·3435	0·3220
2·0	0	0·2385	0·4063	0·4896	0·5107	0·4988	0·4715	0·4399	0·4074	0·3777	0·3500

never penetrates the polar liquid to a great depth, so that instead of the normal concentration of field in the dielectric, making the resonator effectively larger, the liquid actually contains less field than would be in its place in an empty resonator, so that the effective size is decreased.

Table 2. Values of $\mathfrak{K}(Q)$

x ↓ / y →	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
0·5	0·5163	0·5156	0·5131	0·5093	0·5040	0·4973					
0·6	0·5240	0·5233	0·5208	0·5160	0·5106	0·5028	0·4946				
0·7	0·5334	0·5322	0·5294	0·5245	0·5180	0·5098	0·5006	0·4902			
0·8	0·5448	0·5435	0·5402	0·5345	0·5274	0·5181	0·5074	0·4959	0·4838		
0·9	0·5585	0·5570	0·5530	0·5467	0·5380	0·5274	0·5154	0·5026	0·4887	0·4745	
1·0	0·5752	0·5735	0·5687	0·5611	0·5508	0·5384	0·5247	0·5099	0·4945	0·4788	0·4656
1·1	0·5950	0·5929	0·5870	0·5778	0·5656	0·5510	0·5349	0·5173	0·5000	0·4823	0·4660
1·2	0·6188	0·6163	0·6091	0·5975	0·5824	0·5650	0·5459	0·5260	0·5058	0·4855	0·4665
1·3	0·6476	0·6436	0·6349	0·6206	0·6020	0·5806	0·5577	0·5342	0·5111	0·4885	0·4669
1·4	0·6827	0·6785	0·6664	0·6479	0·6243	0·5980	0·5699	0·5424	0·5151	0·4899	0·4661
1·5	0·7267	0·7212	0·7050	0·6802	0·6497	0·6167	0·5824	0·5497	0·5183	0·4897	0·4640
1·6	0·7821	0·7742	0·7515	0·7178	0·6779	0·6357	0·5939	0·5546	0·5187	0·4871	0·4587
1·7	0·8539	0·8422	0·8097	0·7621	0·7086	0·6538	0·6024	0·5560	0·5155	0·4806	0·4500
1·8	0·9503	0·9318	0·8819	0·8149	0·7396	0·6686	0·6055	0·5518	0·5064	0·4687	0·4371
1·9	1·085	1·054	0·9731	0·8695	0·7656	0·6734	0·5990	0·5386	0·4900	0·4510	0·4194
2·0	1·288	1·229	1·0862	0·9231	0·7788	0·6652	0·5781	0·5136	0·4643	0·4260	0·3960

Normally, tube diameters $>0{\cdot}6$ mm. are not used, since the resonance curve is too flat to be measured without errors due to its asymmetry.

Tables 1 and 2 are evaluated from the identity (Watson)

$$J_n(x+y)=\sum_{m=-\infty}^{+\infty} J_m(x)J_{-m}(y).$$

This leads to the following expansions:

$$J_0(x+jy)=J_0(x)I_0(y)-2J_2(x)I_2(y)+2J_4(x)I_4(y)\ldots.$$
$$+j[-2J_1(x)I_1(y)+2J_3(x)I_3(y)\ldots.] \qquad \ldots\ldots(30)$$

and

$$J_1(x+jy)=J_1(x)I_0(y)+J_1(x)I_2(y)-J_3(x)I_2(y)-J_3(x)I_4(y)\ldots.$$
$$+j[J_0(x)I_1(y)-J_2(x)I_1(y)-J_2(x)I_3(y)+J_4(x)I_3(y)\ldots.] \qquad \ldots\ldots(31)$$

in which further terms can be neglected.

In figure 4 is shown the variation of $\mathfrak{Im}(Q)$ with x and y, from which it will be seen that with x constant there is a portion of the curve over which the function hardly varies as y increases. If the capillary tube is selected with such a radius that the function falls on this part of the curve, a large inaccuracy (5%) in the independent value of $\kappa(\alpha y)$ will only result in a small inaccuracy ($\frac{1}{2}$%) in n (αx).

§ 4. EXPERIMENTAL

The resonator, of internal diameter 1·912 cm., is drilled out of a solid brass cylinder, reamed and lapped to an accuracy of $\pm 0{\cdot}02$ mm. It is shown in figure 6. Since a very high Q is unnecessary, the surface need not be silvered; a Q of about 3000 was used throughout. The resonator is fed by two holes half a wavelength apart, 2·7 mm. in diameter, in the short side of a rectangular H_1 waveguide, which forms the centre part of a brass plate terminating the resonator at one end. Power is led from the resonator to the detector through another H_1 waveguide, which is fed from a small hole in the side of the resonator. The axis of this coupling hole makes an angle of 45° with the line of the two feeder holes so as to reduce the effect of unwanted modes. The resonator is tuned by a brass plunger fitting the barrel to 0·5 mm. clearance, and moved by a Moore and Wright micrometer head reading to 0·005 mm. with 2·5 cm. movement. The plunger is backed at its rim by attenuating plastic, which suppresses unwanted modes whose fields on the surface of the resonator are not zero. The glass capillary holding the liquid is a push fit in a hole drilled centrally in the plunger, and slides through two holes drilled centrally in the two short sides of the input waveguide, as the plunger is moved through the resonator. To eliminate dielectric pick-up from the waveguide the glass is

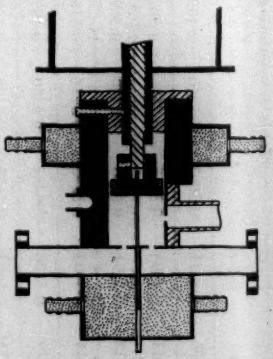


Figure 6. Resonator.

Note.—For convenience the flat for the outlet hole shown normal to the waveguide and not at 45°.

screened by a fine metal tube where it crosses the guide. It is possible for the capillary tube to be central to 0·01 mm. at one end and 0·1 mm. at the other. The resonator is water-jacketed for temperature control both at the waveguide end, and at another point on its barrel. The temperature is read by a thermometer resting in, and making liquid contact with, a hole in the barrel.

The capillaries are drawn from Corning 707 low-loss glass whose dielectric constant is known at the frequency used. They are selected with an internal cross section uniform to one part in two hundred along the portion inside the resonator; this selection is done in the usual way by viewing a length of mercury at various points along the tube under a travelling microscope; the internal diameter is then measured to one part in five hundred by weighing a known length of mercury in the tube, at a known temperature. The walls of the tube should be about 0·05 mm. thick. The outside diameter should be uniform to one part in two hundred along its length, and the tube circular to one part in a hundred. It will be seen from the diagram of the resonator that there is sufficient room to allow an air space and seal at each end of the tube, but usually one end is not completely sealed, so that the water is always at atmospheric pressure.

The apparatus used for transmission and power measurement was similar to that described and illustrated in an earlier paper (Collie, 1946), with the addition of:

1. A piston attenuator which could be inserted into the waveguide instead of the resonator to calibrate the receiver.
2. A second H_{01} resonator connected as a monitor by a one-way feeder for watching frequency stability during a reading. Readings were rejected unless the frequency drift during a set of measurements was less than 50 kcs.
3. Further stabilization of the klystron power packs.
4. Screening round the transmitting tube of the type used in microwave signal generators.

Because of the design of resonator and the tight fitting of the capillary tubes in their plungers, the tubes were not removed for measurement of the Q of the empty resonator; from time to time the entire plunger was removed and replaced by a plain plunger for Q measurement; the change in frequency was measured by comparison with the monitoring resonator which had identical dimensions.

A typical Q curve and the linear plot obtained from it are shown in figure 2. The readings of receiver diode current against resonator tuning are taken for travel in both directions, and are converted to readings of power by use of the calibration curve of the receiver. The plot of inverse of power against square of breadth of Q curve should give a straight line from which the value of α is calculated as previously described. If a straight line is not obtained, it indicates that the resonance is unsatisfactory, or that the calibration of the receiver has changed, and this provides a useful check of each reading.

§ 5. RESULTS

The method described has been tried out on water at 21° c. for which a value of $\kappa = 2{\cdot}83$ is available (unpublished work of the authors). Three series of measurements using three different capillaries were made.

For each measurement, equation (29) has to be solved using the measured value of α and calculating the value of the constant $c = zwu$ from the measured values of r_0, r_1 and r_2 and the known (Penrose, 1946) value of the dielectric constants of the glass.

The following results were obtained:

Table 3. Water at 21° c.

r_0 (mm.)	r_1 (mm.)	$\alpha 10^3$ (cm^{-1})	C	n
0·406	0·495	12·19	10·71	6·28
0·525	0·590	39·90	6·51	6·30
0·500	0·610	33·40	7·00	6·35

$r_2 = 0{\cdot}9560$ cm.,
$\epsilon' = 4{\cdot}00$,
$\epsilon'' < 0{\cdot}01$. For Corning 707 low-loss glass.

§6. DISCUSSION

Over the course of two years we have tried many methods of measuring the dielectric constant of water. The method described is the only one we have found for n at 1·25 cm. which is capable of $\pm 1\%$ accuracy, and a series of systematic measurements of the variation of n with temperature is being planned.

There is no doubt that if this work were being done again we should adopt the M.I.T. (Pound, 1946) system of frequency stabilization, which locks the oscillator on to a resonant cavity.

§7. ACKNOWLEDGMENTS

The work described in this paper has been carried out on behalf of the Director of Physical Research, Admiralty, and the authors wish to record their thanks for permission to publish. They thank Lord Cherwell for his continued interest and encouragement.

The suggestion of using an H_{01} resonator arose in a talk with Dr. L. A. P. Speirs, with whom we had many helpful discussions.

We are grateful to Mr. H. Ashcroft for the method of evaluating equations (30) and (31).

REFERENCES

Bethe, 1942, " Perturbation theory of cavity resonators ", *M.I.T. Report.*
Collie, 1944, *Proc. Phys. Soc.*, **56**, 255.
Collie, Hasted and Ritson, 1946, *Trans. Faraday Soc.* (Discussion on Dielectrics).
Penrose, 1946, *Trans. Faraday Soc.* (Discussion on Dielectrics).
Pincherle, 1944, *Phys. Rev.*, **66**, 118.
Pound, R. V., 1946, *Rev. Sci. Instrum.*, **17**, 490.
Watson, *Theory of Bessel functions* (Cambridge University Press), p.30.
Willis Jackson, 1941, *Ministry of Supply*, R. 287/Gen/35.

Pressure Broadening of the Inversion Spectrum of Ammonia: Part II—Disturbance of Thermal Equilibrium at Low Pressures

By B. BLEANEY and R. P. PENROSE
Clarendon Laboratory, Oxford

MS. received 14 *March* 1947

ABSTRACT. The absorption at the centre of the line (3, 3) of the centimetre wavelength inversion spectrum of ammonia gas has been measured at pressures between 1·5 mm. and 0·01 mm. Hg. The absorption coefficient is constant at the higher pressures, as would be expected for a single line whose width is determined solely by pressure broadening. At the lower pressures the absorption coefficient falls by an amount dependent on the energy density in the resonator; this is due to the disturbance of thermal equilibrium in the gas through the absorption of energy from the radio-frequency field. A theory of the effect is developed from which the thermal relaxation time can be calculated by comparison with the experimental values of the absorption coefficient. It is found that this time is approximately $1{\cdot}7_5$ times greater than the mean time between collisions which interrupt the absorption (determined from the line breadth constant). This indicates that not every such collision is effective in restoring thermal equilibrium.

§1. INTRODUCTION

The measurement of the absorption in ammonia gas at pressures of 1 cm. Hg to 60 cm. Hg has been described in a previous paper (Bleaney and Penrose, 1947 a). The observed values at pressures of 60 cm. and 10 cm. Hg were compared theoretically on the basis of the earlier analysis of the spectrum at pressures of about 1 mm. Hg (Bleaney and Penrose 1946 a, 1947 b). The observed and calculated values of the absorption at 10 cm. pressure were in close agreement over the whole range of the measurements (0·63–0·92 cm^{-1}), showing that the widths of the lines varied accurately at the pressure between 0·5 mm. Hg and 10 cm. Hg.

At pressures below 1 mm. Hg, the widths of most of the lines in this spectrum are small compared with the separation of the lines, and it is therefore possible to examine the lines individually. The width of the lines is still due almost entirely to collision broadening, since at a pressure of 0·5 mm. Hg the line breadth constants (equal to half the width of the line at half intensity) are found to lie between $2 . 10^{-4}$ and $5 . 10^{-4}$ cm^{-1}. Of other sources of broadening, that due to the Doppler effect was expected to be the most important; the line breadth constant associated with it is approximately 10^{-6} cm^{-1}, and its effect should not therefore become appreciable until pressures of about 10^{-3} mm. Hg are reached. Thus it was expected that a range of pressures varying by a factor of nearly 1000 should be available for the study of the effects of collision broadening on single lines.

In a region where the absorption is all due to a single spectral line whose width is determined entirely by collision broadening, the absorption at the centre of the line should be independent of the pressure. This is evident from the formula

$$\frac{\alpha}{\bar{\nu}^2} = \frac{4\pi^2 |\mu_{mn}|^2}{3kT} \frac{N_{mn}}{\Delta\bar{\nu}} \qquad \ldots\ldots(1)$$

where α = absorption coefficient per cm. of path,

N_{mn} = number of molecules per cc. in the lower and upper energy levels of the transition,

$\Delta\bar{\nu}$ = line breadth constant = half the width of the line at half intensity;

since both N_{mn} and $\Delta\bar{\nu}$ vary linearly with pressure, and the remaining constants are independent of the pressure. Since measurement of the absorption coefficient at the centre of the line is simpler and more accurate than a direct measurement of the width, it was decided to use this measurement as a test of the validity of equation (1). The line (3, 3) was chosen as the strongest and most convenient line for this purpose. Contrary to expectations, it was found that the absorption coefficient began to decrease when the pressure was about a tenth of a millimetre, and fell rapidly as the pressure was reduced still further. Investigation of the other lines of the spectrum showed that a similar effect occurred with all of them. The absorption at their centres began to fall at pressures of about 0·1 mm. and continued to decline as the pressure was reduced.

This fall in the absorption coefficient might be due to some other form of broadening of the line, which is independent of the pressure; the line breadth constant associated with this broadening must be of the order of 10^{-4} cm.$^{-1}$ to explain the observations. The following possibilities were considered:

(*a*) The Doppler width

Owing to the velocities of the molecules of a gas, a spectral line has a Doppler width given by the expression

$$\Delta\bar{\nu} = \left(\frac{\bar{\nu}}{c}\right)\sqrt{\frac{2RT}{M}\log_e 2}. \quad (\text{cm.}^{-1})$$

For ammonia, the value of $\Delta\bar{\nu}$ given by this is $1{\cdot}2 \cdot 10^{-6}$ cm.$^{-1}$

(*b*) Interruption of the absorption owing to collisions with the walls.

In a container of volume V and wall area A, a molecule of mean velocity v will make a collision with the wall $\frac{1}{4}Av/V$ times a second. For the resonator used in this experiment, this will cause a broadening of the line of approximately

$$\Delta\bar{\nu} = \tfrac{1}{4}Av/2\pi cV,$$
$$\doteq 3 \cdot 10^{-7}\,\text{cm.}^{-1}$$

(*c*) The natural line breadth, due to spontaneous emission.

The line breadth constant associated with spontaneous emission is

$$\Delta\bar{\nu} = \frac{32\pi^3\nu^3|\mu_{mn}|^2}{3hc}. \quad (\text{cm.}^{-1}) \qquad \ldots\ldots(2)$$

For wavelengths of about 1 cm. this gives a value of $\Delta\bar{\nu}$ of about 10^{-18} cm.$^{-1}$ In Einstein's theory of spontaneous emission, the density of radiant energy is assumed to be given by Planck's law; Purcell (1946) has, however, pointed out that with a resonator is associated one oscillator in a frequency range ν/Q, rather than $8\pi\nu^2/c^3$ per unit volume per unit band width, as with Planck radiation theory. Thus with a resonator Q of 10 000 the line breadth constant will be about 10^{-15} cm.$^{-1}$

(*d*) The natural line breadth due to stimulated emission and absorption.

A molecule in thermal equilibrium with its surroundings will undergo transitions owing to interaction with the radiation field. In the optical region the number of such transitions is small compared with the number of transitions due

to spontaneous emission, but the reverse is true at radio frequencies. The number n of transitions per second that a molecule undergoes in equilibrium in thermal radiation of energy density ρ is

$$n = \frac{8\pi^3 |\mu_{mn}|^2}{3h^2} \rho. \quad (\text{sec.}^{-1}) \qquad \ldots\ldots(3)$$

At these frequencies where $ch\bar{\nu} \ll kT$, this becomes

$$n = \frac{64\pi^4 |\mu_{mn}|^2 \bar{\nu}^2 kT}{3h^2 c} \quad (\text{sec.}^{-1}) \qquad \ldots\ldots(3\,a)$$

whence

$$\Delta\bar{\nu} = n/2\pi c = \frac{32\pi^3}{3h^2c^2} |\mu_{mn}|^2 \bar{\nu}^2 kT. \quad (\text{cm.}^{-1}) \qquad \ldots\ldots(3\,b)$$

This formula gives the width of a radio-frequency spectral line as observed in emission; for the inversion lines of ammonia this $\Delta\bar{\nu}$ is about 10^{-16} cm.$^{-1}$, or, if the emission excites a resonator tuned to the frequency of the line, $\Delta\bar{\nu}$ will again be about 1000 times greater. In absorption measurements the resonator is excited by radiation of an intensity immensely greater than that corresponding to room temperature, and the width of the line in absorption will be correspondingly greater. This point will be discussed later.

(*e*) Interruption of the inversion by a transition to another rotational level. Since the inversion frequency of the molecule depends strongly on the rotational state of the molecule, the absorption of radiation corresponding to a transition between the two levels of the inversion doublet will be interrupted if the molecule jumps to another rotational level. Transitions between the rotational levels will be caused by spontaneous emission, and by absorption and stimulated emission due to the thermal radiation. From equations (2) and (3) it can be seen that transitions between the rotational levels are considerably more frequent than similar transitions between the inversion sub-levels, since the energy density of black-body radiation is much greater at the higher frequencies of the rotational spectrum of ammonia. Calculation shows that the number of quanta absorbed per second by a molecule in the $J = 3$ level causing a transition to the $J = 4$ level is about 10 per second; the number of spontaneous and stimulated emissions causing transitions to the $J = 2$ level are each of the same order. Hence the associated line breadth constant is about 10^{-10} cm.$^{-1}$

It is apparent that none of these phenomena is sufficient to cause a broadening comparable even with that due to the Doppler effect, except that due to the stimulated transitions considered under (*d*) above. This phenomenon cannot be considered as a simple broadening, but as a disturbance of thermal equilibrium in the gas. The absorption of energy from the radio-frequency field tends to equalize the populations of the two energy levels of the inversion doublet; this process is counteracted by the action of the collisions between the molecules, which tends to restore thermal equilibrium. As the pressure is reduced, the time between collisions gets longer, and the "relaxation time" for the return to thermal equilibrium increases. When the collision frequency becomes so low that the probability of making a transition in the interval between collisions approaches unity, the

average populations of the energy levels will no longer correspond to that appropriate to the "temperature" of the gas, but to some higher temperature. The intensity of the absorption will then decrease.

That this effect is of the correct order of magnitude to explain the observed diminution in the absorption coefficient is shown by the following calculation. If N_{JK} is the number of ammonia molecules per cc. in the rotation level characterized by the quantum numbers J, K, then the excess number in the lower level of the inversion doublet is $\frac{1}{2}N_{JK}(h\nu_0/kT)$, where $h\nu_0$ is the energy separation of the two levels of the doublet. To equalize the populations of the two levels, half of these must be transferred to the upper level, which requires an energy of $h\nu_0 \cdot \frac{1}{4}N_{JK} \cdot (h\nu_0/kT)$. This is the maximum energy that the gas can absorb from the radiation in a time (τ) equal to the mean interval between collisions.* Hence the power dissipated per cc. is

$$W = \tfrac{1}{4}N_{JK} \cdot (h\nu_0)^2/kT. \qquad \ldots\ldots(4)$$

For the line $(J, K) = (3, 3)$ at a pressure of 0·1 mm. Hg $(1/\tau) = 1{\cdot}8 \cdot 10^7 \text{ sec.}^{-1}$, while $N_{JK} = 2{\cdot}3 \cdot 10^{13}$; hence $W = 200$ ergs/cc./sec. Now the power dissipated in the resonator in these experiments is of the order of 1/10 milliwatt, or 1000 ergs/sec. At the centre of a strong line such as (3, 3) some 2/3 of this power would be absorbed by the ammonia, in a volume (allowing for the distribution of electric field in the resonator) of less than 2 cc., if the absorption has its normal intensity. Since this would require an absorption of energy by the ammonia greater than that given by equation (4), it is obvious that the strength of the absorption must be less than that corresponding to the maintenance of thermal equilibrium.

The treatment of the last paragraph leads to a simple expression of the condition that must be fulfilled if the true value of the absorption in a gas is to be measured. This condition is that the power absorbed by the gas must be small compared with the power given by equation (4). This equation contains two points of interest:

(*a*) W depends on the square of the resonant frequency of spectral line. This arises because the energy absorbed in a transition varies with ν and so does the difference of population of the upper and lower levels of the transition.

(*b*) W varies with the square of the pressure in the gas. This follows from the fact that both N_{JK} and $(1/\tau)$ vary linearly with the pressure.

The correctness of this explanation of the fall in the absorption coefficient at pressures of about 1/10 mm. Hg could quickly be verified experimentally by measurement with different power levels in the resonator. From (*b*) of the last paragraph it would be expected that the drop in the absorption coefficient would be the same under conditions such that the product (power dissipated in the resonator times square of the pressure of ammonia) is constant. This was found to be the case.

The phenomenon of the disturbance of thermal equilibrium in radio-frequency spectroscopy owing to the absorption of energy from the radiation was suggested and observed in experiments on the nuclear magnetic resonance by Purcell,

* For the purposes of this calculation it is assumed that the thermal relaxation time is identical with the mean time between collisions which interrupt the absorption of radiation.

Torrey and Pound (1946), and Bloch, Hansen and Packard (1946). In these experiments with liquid or solid samples, the thermal relaxation time is unknown, even as to order of magnitude, and a main interest lies in its determination by means of the decrease in the apparent absorption coefficient as the energy density of the radiation is increased. In the case of ammonia, the thermal relaxation time should be of the same order as the mean time between collisions* which is accurately known from the width of the spectral line, measured at low energy density. The opportunity arises, however, of determining whether every collision which is effective in interrupting the absorption of radiation is also effective in tending to restore thermal equilibrium. This may be accomplished by measurement of the thermal relaxation time by observation of the variation of the absorption coefficient with pressure at various power levels, which is more convenient experimentally than the complementary method of observing the decrease in the apparent absorption at a given pressure as the power level is increased. In the following sections the theory of the method is developed, and the predicted results are compared with those obtained experimentally by measurements at the centre of the line (3, 3). From this comparison the ratio of the thermal relaxation time to the mean time between collisions is obtained.

§ 2. THEORY

Since the strength of the absorption depends only on the *difference* between the number of molecules occupying the lower and the number occupying the higher of the two energy levels between which transitions are taking place, it is sufficient to work in terms of this difference, rather than the actual numbers themselves. Let n_0 be the excess number of molecules in the lower state corresponding to the actual temperature of the gas, and let n be the actual excess when radio-frequency power is applied. Then in the equilibrium condition the rate at which the r.f. power is causing n to decrease must be balanced by the action of collisions which tend to restore n to the value n_0 corresponding to thermal equilibrium. If β is the number of transitions per second per molecule caused by the radiation, then the rate of change of n is $(-2\beta n)$, since the transfer of one molecule from the lower to the upper state reduces the excess by two. The rate of change of n due to collisions may be written (cf. Fröhlich, 1946) as $(n_0-n)/\tau_e$, where τ_e is the mean time between collisions that are effective in tending to restore thermal equilibrium. Hence we have

$$\frac{dn}{dt} = -2\beta n + (n_0 - n)/\tau_e. \qquad \ldots\ldots(5)$$

In the steady state $dn/dt = 0$; hence

$$n/n_0 = 1/(1+2\beta\tau_e) \qquad \ldots\ldots(5a)$$

and this equals the ratio of the observed absorption coefficient (α) to the true absorption coefficient (α_0) which would be observed if the energy density of the radiation were infinitely small.

Since the ratio n/n_0 is only appreciably different from unity for very low pressures, where the spectral lines are narrow, it is sufficiently accurate in the

* In this paper, where the word "collisions" occurs unqualified, it is to be understood as referring to collisions which produce interruption of the absorption of radiation. This restriction is introduced to avoid constant repetition of this phrase.

evaluation of β to neglect the second-order term introduced by Van Vleck and Weisskopf (1945) and Fröhlich (1946) into the collision broadening theory of Lorentz. Then at a place where the electric field of the radiation is $E\cos 2\pi\nu t$ the chance of a molecule making a transition between two energy levels W_m and W_n in a time t is

$$(a_m^* a_m) = |\mu_{xmn}|^2 E^2 \frac{\sin^2\left\{\frac{\pi}{\hbar}(W_m - W_n - \hbar\nu)t\right\}}{(W_m - W_n - \hbar\nu)^2} \qquad \ldots\ldots(6)$$

(cf. Pauling and Wilson, *Introduction to Quantum Mechanics*, Chap. XI, equation (40–12)). Here

$$\mu_{xmn} = \int \psi_m^{0\,*} \mu_x \psi_n^0 \, d\sigma$$

where μ_x is the component of the electric dipole moment in the direction of the electric field of the radiation. Since the dipoles are randomly oriented,

$$(\mu_{xmn})^2 = (\mu_{ymn})^2 = (\mu_{zmn})^2 = \tfrac{1}{3}(\mu_{mn})^2.$$

In the measurements to be described the frequency of the applied radiation corresponded to the centre of the spectral line; i.e. $\hbar\nu = W_m - W_n$ and (6) becomes

$$(a_m^* a_m) = \frac{\pi^2}{\hbar^2}\frac{|\mu_{mn}|^2}{3} E^2 t^2. \qquad \ldots\ldots(6a)$$

The value of t in this equation is that corresponding to the actual time between successive collisions; since this time is random, an average must be taken over all possible values. For present purposes it is sufficient to assume an exponential distribution $(1/\tau)\,e^{-t/\tau}$ of the values of t. Then, since $\overline{t^2} = 2\tau^2$, (6 *a*) becomes

$$\overline{(a_m^* a_m)} = \frac{2\pi^2}{\hbar^2}\frac{|\mu_{mn}|^2}{3} E^2 \tau^2 \qquad \ldots\ldots(6b)$$

and the β of equation (5) equals $\overline{(a_m^* a_m)}/\tau$.

The appropriate value of τ in equation (6 *b*) is the mean time between collisions which are effective in interrupting the absorption of radiation, and τ is therefore related to the line breadth constant for collision broadening. The value of τ may therefore be found by observation of the width of the spectral line under conditions where broadening due to causes other than collisions between the molecules is negligible.

In applying equation (6 *b*) to a case where the absorption of energy from the radiation is so great that thermal equilibrium is disturbed, it is necessary to examine its validity under these conditions. The expression for the transition probability (equation (6)) is derived by means of first-order perturbation theory, where the assumption is made that the chance of a transition in the time t is very small. It is obvious from equation (5 *a*), however, that n/n_0 will only be materially smaller than unity when the term $2\beta\tau_e$ is no longer small. If τ_e is approximately equal to τ, the value of $\overline{(a_m^* a_m)}$ must be of the same order as $\beta\tau_e$; that is, unity or more. Thus equation (6 *b*) is here being applied under conditions where its method of derivation is no longer justified, if $\bar{n}$ is materially less than n_0. However, in the absence of an exact theory* it is interesting to examine whether the variation of the absorption

* Professor M. H. L. Pryce informs us that preliminary work on a theory which avoids these limitations indicates that the formula (equation (7)) for (α/α_0) is valid even when the apparent absorption coefficient is materially smaller than the true value.

coefficient observed experimentally can be adequately represented by formulae deduced in this way.

If the absorption is observed using a plane wave traversing the unbounded medium, equation (5) becomes simply

$$\alpha/\alpha_0 = n/n_0 = 1\Big/\left(1+\frac{4\pi^2}{3\hbar^2}|\mu_{mn}|^2 E^2 \tau\tau_e\right) \qquad \ldots\ldots(7)$$

$$= 1\Big/\left(1+\frac{32\pi^3}{3\hbar^2 c}|\mu_{mn}|^2 W\tau\tau_e\right) \qquad \ldots\ldots(7a)$$

where $W = cE^2/8\pi$ = energy transmitted across unit area per second.

In a waveguide, or cavity resonator, the effect is more complicated owing to the inhomogeneous distribution of the electric field. Most of the absorption occurs at a place where the electric field is strongest, and at this point the value of n/n_0 will fall below unity by the greatest amount. At other places the disturbance of thermal equilibrium owing to the effect of the radiation will be less, and the values of n/n_0 will be closer to unity. Clearly the observed value of α/α_0 will be that corresponding to an average value of n/n_0, obtained by evaluation of the integral

$$\alpha/\alpha_0 = \overline{n/n_0} = \frac{\int\{E^2/(1+2\beta\tau_e)\}dV}{\int E^2 dV}, \qquad \ldots\ldots(8)$$

the integral being taken over the whole volume in which radiation is acting on the gas. An explicit expression for this integral cannot be given for the general case, but it may be shown that for a given shape of cavity and mode of excitation its value depends only on the non-dimensional parameter

$$a' = (4\pi^2/3\hbar^2) \cdot |\mu_{mn}|^2 \cdot \tau\tau_e \cdot (gW'/V),$$

where W' = total energy stored in cavity,

V = total volume of cavity,

g = dimensionless constant calculated from the configuration of the electric field.

In these experiments a cylindrical cavity in the H_0 mode was used, and the relation between (α/α_0) and a' is shown graphically in figure 1. The evaluation of the integral in equation (8) for this case is given in Appendix A.

For the comparison of this theory with the experimental results the values of the variables in the parameter (a') are required. For the inversion lines of ammonia the matrix element for a rotational state denoted by the quantum numbers (J, K) is

$$|\mu_{JK}|^2 = \mu^2 \cdot K^2/(J^2+J).$$

The remaining quantities can only be determined by experiment. The mean time

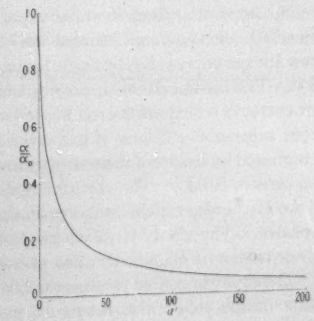


Figure 1. Fractional drop in absorption (α/α_0) as a function of power level in an H_0 cylindrical resonator.

between collisions which interrupt the absorption of radiation (τ) is found from the line breadth constant ($\Delta\bar{\nu}$) measured under conditions where only collisions are significant in broadening the line, using the relation

$$\tau = 1/(2\pi\Delta\bar{\nu}c).$$

The total energy W' stored in the resonator can be found by measuring the power dissipated in the resonator and the circuit magnification factor Q; the volume V is known from the dimensions of the cavity.

There remains the thermal relaxation time τ_e which cannot be determined independently. Since, however, the thermal relaxation is due to collisions between the molecules, one would expect that τ_e should vary inversely with the pressure in the same way as the collision time τ. It is therefore convenient to work in terms of the ratio of these two times (τ_e/τ) as a parameter whose value should be independent of the pressure. The value of a' may then be written

$$a' = \frac{|\mu_{JK}|^2}{3(hc)^2}\frac{(\tau_e/\tau)}{\Delta\bar{\nu}^2}\left(\frac{gW'}{V}\right)$$

where only the ratio (τ_e/τ) is undetermined.

The development of the theory has thus been carried as far as possible, and its results can be subjected to experimental test. The purpose of such a test is two-fold:

(1) To examine whether the formulae derived by this theory are applicable when thermal equilibrium is seriously disturbed, and the ratio of the apparent to the true absorption coefficient is very much less than unity.

(2) To determine the ratio of the thermal relaxation time τ_e to the mean time between collisions τ.

If it is found possible to represent the whole range of the experimental results by curves calculated from the theory with a single value of the undetermined parameter (τ_e/τ), it seems justifiable to assume that the theory is satisfactory and that the results yield a value of the thermal relaxation time.

§3. EXPERIMENTAL APPARATUS

The apparatus used in these experiments was the same as that previously described (Bleaney and Penrose 1947 b) for measurements on the spectrum of ammonia gas at pressures of about 1 mm. Hg, and only a brief outline need be given here. The lay-out of the apparatus can be seen from the photograph (figure 2). Power from a reflexion klystron A is fed into a rectangular waveguide, its magnitude being adjusted by means of the variable attenuator C. Its wavelength can be determined by means of the sharply tuned wavemeter E with an accuracy of one or two parts in 10000. The experimental chamber H is a cylindrical cavity resonant in the H_0 mode, which can be evacuated or filled with ammonia at any desired pressure. The cavity is tuned by a plunger driven by a micrometer head M, whose motion is magnified, when necessary, by the mirror, lamp and scale.

The cavity is excited by power fed from the waveguide through two small holes in the narrow side of the rectangular guide. The waveguide is terminated by an attenuator J, followed by a silicon-tungsten crystal rectifier (not visible) which acts as a monitor. To measure the power flowing along the guide, this can be replaced by a thermistor bolometer. Resonance in the cavity is detected by

another crystal rectifier K, preceded by an attenuator L, coupled to the cavity through a small hole in the side of H. The rectified current from this crystal is proportional to the energy density in the cavity, and the strength of the absorption of the ammonia is measured by means of the change in the energy density in the cavity on admitting the ammonia. If δ_0, δ_1 are the crystal current readings before and after admitting the ammonia, then the absorption coefficient per cm., α, may be calculated from the relation

$$\alpha = \frac{2\pi}{\lambda Q_0}\left\{\sqrt{\frac{\delta_0}{\delta_1}} - 1\right\}, \qquad \ldots\ldots(12)$$

where λ is the wave-length of the radiation in free-space and Q_0 is the circuit magnification factor of the empty resonator. Q_0 is determined from the movement of the micrometer head required to detune the resonator to the half-power points, and can be measured with an accuracy of a few per cent.

Measurements

To test the theory of the disturbance of thermal equilibrium developed above, the procedure adopted was to measure the apparent absorption at the centre of the line (3, 3) at pressures from 1·5 mm. Hg down to a few hundredths of a mm., using two different power levels. This line is one of the strongest in the ammonia spectrum, and at these pressures the absorption due to the tails of neighbouring lines is very small. The centre of this line lies at 0·7964 cm.$^{-1}$, and can be determined very accurately by locating the maximum of the absorption at a pressure of about 0·1 mm. The reading of the micrometer head attached to the resonant cavity corresponding to this maximum was observed in a preliminary experiment, and the subsequent measurements were all made with the frequency of the oscillator adjusted to be within 0·00002 cm.$^{-1}$ of the exact centre of the line.

The amount of r.f. power entering the cavity cannot be measured directly, but may be deduced from measurements of the power reaching a detector placed at the end of the input waveguide when the cavity is first detuned from, and then tuned to, resonance. A simple calculation based on the assumption that both the generator and the detector are matched to the waveguide shows that if the drop in the detector reading is a small fraction r of its initial reading, then the power entering the resonator is a fraction $r/(1+r)$ of that incident on the detector before tuning the cavity. The power flowing along the main waveguide was measured by a bolometer whose sensitive element was a thermistor. On adjusting the cavity to resonance the reading of the bolometer fell by 13% and the power entering the cavity is therefore a little under 12% of the incident power. The remaining 1% is reflected back to the generator owing to the slight mismatch caused by the impedance reflected into the guide when the cavity is tuned.

At the lowest power level the thermistor was not sufficiently sensitive to measure accurately the drop in power on tuning the cavity as the power incident on the thermistor could only be determined to within 10%. To verify that the fraction of this power entering the cavity was the same as at the higher level, the thermistor was removed, and the attenuator J followed by a crystal detector was replaced. It was found that the drop in the rectified current from the crystal on tuning the cavity was in good agreement with that obtained with the thermistor at the higher power level.

These methods were used to determine the power dissipated, and hence the energy density in the resonant cavity when it was evacuated. When ammonia is admitted the Q of the cavity drops owing to the absorption of energy by the gas, and less power is drawn by the cavity from the main waveguide. The energy stored in the cavity changes, therefore, with the square of Q. It is not necessary to determine the Q and the power entering the cavity at each pressure of ammonia, however, since the change in the energy density in the cavity is measured directly by the change in the power registered by the crystal detector K. From the latter the absorption coefficient of the ammonia is also calculated by equation (12).

The power output from the oscillator to the waveguide remained constant to within 2% during the experiments. The procedure was therefore as follows: After the measurement of the power entering the evacuated resonator the rectified current of about 1 μa. from the crystal K was observed by a sensitive galvanometer. Ammonia was then admitted to the cavity, its pressure being measured by the McLeod gauge shown in figure 2, and the new reading of the galvanometer observed. The cavity was evacuated and the procedure repeated with different pressures of ammonia until sufficient points were obtained to delineate a curve of absorption coefficient against pressure.

§4. RESULTS

Measurements were made at two power levels differing by a factor of 70. At the higher level the power in the waveguide was 12·3 mw. at the coupling to the resonant cavity; this was the maximum that could be obtained using sufficient attenuation in the guide to prevent errors due to change of oscillator frequency on tuning the resonator. The power entering the resonant cavity when evacuated was 1·4 mw.

In the second series of experiments, the power was reduced to as low a value as possible, the limit being set by the sensitivity of the detector. The power level in the waveguide was 0·16 mw., and the power entering the resonator was approximately 20 μw.

The resonator was used at its second resonance and was 1·82 cm. long; its radius was $1{\cdot}05_4$ cm. The circuit magnification factor Q_0 of the empty resonator was measured and found to be 8500.

The change of the absorption coefficient with pressure at the two power levels is shown in figure 3, where the absorption is expressed both per cm. of path and in decibels per kilometre. It will be seen that the drop in the apparent absorption as the pressure is reduced occurs at much greater pressures when the power level is higher. In fact the pressures in the two curves at which the absorption has fallen to a given value are very closely in the ratio of $\sqrt{70}:1$ i. e. the square root of the ratio of the two powers, as would be expected if the phenomenon is due to disturbance of thermal equilibrium. At pressures above 0·2 mm. Hg the absorption coefficient is very nearly constant when the lower power level is used, showing that contributions to the width of the line from causes other than collision broadening are negligible. The experimental values of the absorption coefficients show a slight increase at the higher pressures, due to the contributions from neighbouring lines. The additional absorption from this cause has been calculated, and subtracted from the experimental points before plotting them in figure 3.

To compare these results with the variation of absorption with pressure and power level predicted by the theory developed earlier in this paper, the width of the spectral line is required, measured at low energy density so that thermal equilibrium is not disturbed. As reported previously (Bleaney and Penrose 1947 b) the width is found from the shape of the line at a pressure of 0·5 mm. Hg, determined by measurement of the absorption coefficient at a sufficient number of frequencies to delineate the line. These measurements are shown in figure 4, where the absorption coefficient is plotted against the wavelength in the resonant cavity, small

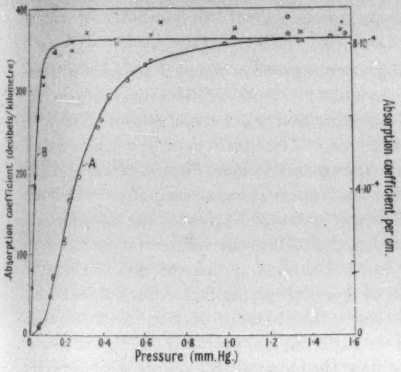


Figure 3. Variation of absorption with pressure at the centre of the line $(J, K)=(3, 3)$ for two different power levels.

A. Power entering evacuated cavity = 12·3 milliwatts.
○ Experimental points.
— Theoretical curve, assuming $(\tau_e/\tau)=1{\cdot}75$.

B. Power entering evacuated cavity = 20 microwatts.
× Experimental points.
— Theoretical curve, assuming $(\tau_e/\tau)=1{\cdot}75$.

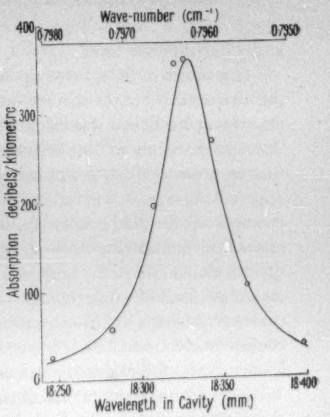


Figure 4. The line $(J, K)=(3, 3)$ at a pressure of 0·5 mm. Hg.

changes in which are proportional to changes in the wave-number. The full line shows the theoretical absorption curve, calculated from the formula

$$\alpha_0=\frac{4\pi^2|\mu_{JK}|^2\bar{\nu}^2N_{JK}}{3\boldsymbol{k}T}\left\{\frac{\Delta\bar{\nu}}{\Delta\bar{\nu}^2+(\bar{\nu}_0-\bar{\nu})^2}\right\}$$

where

$$|\mu_{JK}|^2=|\mu_{3,3}|^2=\tfrac{3}{4}(1{\cdot}44\,.\,10^{-18})^2 \text{ e.s.u.}$$

The value $4{\cdot}5\,.\,10^{-4}$ cm.$^{-1}$ for the line breadth constant $\Delta\bar{\nu}$ has been taken as giving the best fit with the experimental points. The agreement between the experimental and theoretical values is very close, and the error in the value of $\Delta\bar{\nu}$ cannot be more than 1 or 2%. Since $\Delta\bar{\nu}$ varies linearly with the pressure, one may write

$$\Delta\bar{\nu}=9{\cdot}0\,.\,10^{-4}p_{\text{mm}}. \quad (\text{cm.}^{-1})$$

An expression relating the constant a' of equation (11) with the experimental data can now be found. The energy stored in the resonant cavity is $W'=QW/\omega$. Here the measured quantities are not W and Q but W_0, the power dissipated in the

empty resonator, whose circuit magnification is Q_0. When ammonia is admitted so that the rectified current from the crystal detector K falls from δ_0 for the evacuated resonator to δ_1, the value of W' is

$$W' = (Q_0 W_0/\omega)(\delta_1/\delta_0).$$

From these equations, and data given earlier in this section, the value of a' is found to be

$$a' = (0{\cdot}63/p^2_{\mathrm{mm}})(\delta_1/\delta_0)$$

at the higher power level, and

$$a' = (0{\cdot}0090/p^2_{\mathrm{mm}})(\delta_1/\delta_0)$$

at the lower power level.

The values of (δ_1/δ_0) corresponding to each point in figure 3 are known from the measurement of the absorption. As a first attempt, the theoretical drop in the absorption coefficient was calculated assuming that the thermal relaxation time τ_e was equal to the mean time between collisions. The curves of (α, p) constructed in this way were found to be similar to the experimental curves, but the decrease of the apparent absorption was rather less than that observed experimentally. For both power levels the discrepancy was such that it appeared likely that the introduction of a simple multiplying factor of rather less than 2 into the values of a' would shift the calculated curves for both power levels into good agreement with the experimental points. To find the best value of this multiplying factor, the values of the constant a' which would be required to give the observed drop in the absorption coefficient were calculated for each of the experimental points at the higher power level. It was found that the values of a' for the most reliable points were between 1·7 and 1·8 times greater than those calculated on the assumption $(\tau_e/\tau) = 1$. The theoretical curves for both power levels were therefore recalculated assuming $(\tau_e/\tau) = 1{\cdot}7_5$ and are shown as full lines in figure 3. It will be seen that the experimental results are in good agreement with the theoretical predictions made on this basis.

§ 5. DISCUSSION

Although the absorption coefficient has fallen at the lower pressures to about one-twentieth of the value which would be observed if the energy density of the radiation were sufficiently small, the theory is still successful in representing the experimental results. It may therefore be inferred that the formula (7) for the apparent diminution in the absorption is valid even under these conditions where its method of derivation cannot be justified. The success of the theory makes it permissible to discuss the significance of the thermal relaxation time found by means of these experiments.

A few simple considerations indicate that the value of the ratio of the thermal relaxation time (τ_e) to the mean time (τ) between collisions which are effective in interrupting the absorption of radiation should lie between fairly narrow limits. Since it seems impossible to conceive of a collision mechanism which is effective in producing thermal relaxation without interrupting the absorption of radiation, it is reasonable to suppose that (τ_e/τ) cannot be less than unity. On the other hand a "kinetic theory collision", i. e. a collision in which there is an appreciable transfer of momentum between the molecules, must surely be effective in producing thermal relaxation. This leads to an upper limit for (τ_e/τ). The value found for

(τ_e/τ) in these experiments by no means approaches this upper limit, however. Viscosity measurements show that an appreciable transfer of momentum between ammonia molecules only occurs when they approach to within a distance of about 3 A., whereas the value of the line breadth constant for this absorption shows that the interaction of the molecule with the radiation field is interrupted by the approach of another molecule to a distance of about 14 A.

Consideration of the experimental errors suggests that the value $1{\cdot}7_5$ of the ratio (τ_e/τ) assumed in fitting the theoretical to the experimental values is probably correct to within $\pm 10\%$. The distance between the molecules at which the interaction is sufficient for the purposes of thermal equilibrium is thus about $14/\sqrt{1{\cdot}75} = 11$ A., which is still considerably greater than the kinetic-theory diameter of the molecule. It must be remembered, however, that the absorption only effects the distribution of molecules amongst the two energy levels of the inversion doublet whose separation is very small ($0{\cdot}8\,\mathrm{cm.}^{-1}$). The restoration of the correct thermal distribution may therefore be effected by the interaction of two molecules at a distance much greater than that required for the transfer of momentum.

A lengthy discussion of the significance of the difference between the thermal relaxation time and the mean time between collisions which interrupt the absorption of radiation would seem to be premature. The theory of collision broadening of Van Vleck and Weisskopf is based on the concept of a classical harmonic oscillator whose transient amplitude and phase after a collision are distributed in accordance with a Boltzmann function. Thus every collision is effective in restoring thermal equilibrium. On the other hand, the concept of the disturbance of thermal equilibrium is based on the existence of discrete energy levels, and the theory is essentially quantum-mechanical, as is obvious from the presence of Planck's constant in the formulae. It is thus very difficult to discuss the phenomenon from the point of view of the theory of Van Vleck and Weisskopf, and no analogous theory has yet been developed on quantum-mechanical principles. It seems pertinent to enquire, however, what meaning can be attached to a "collision" on such a theory. The absorption of energy from the radiation tends to equalize the numbers of molecules in the upper and lower levels between which transitions are taking place, and thermal relaxation tends to restore the unequal distribution appropriate to the temperature of the gas. That is, thermal relaxation collisions cause a net transfer of molecules from the upper to the lower level. The fact that τ is less than τ_e shows, however, that a large proportion of the collisions which interrupt the absorption are not effective in causing such a transfer. The fact that they cause a broadening of the spectral line shows that they must effect the wave function of the molecule in some way, even though they leave the molecule in the same energy level. It may prove necessary, therefore, to consider the phase of the wave function (which is ordinarily neglected), and to attribute to a collision a random change of phase in the wave function similar to the change of phase of the harmonic oscillator in the classical theory. Full elucidation of this point must await the proper formulation of a wave-mechanical theory of collision broadening.

§6. CONCLUSION

Apart from its intrinsic interest in the study of collision broadening, the phenomenon described in this paper is important in connection with the limitation of the

resolution of close spectral lines which it causes. This limitation arises in a two-fold manner; the absorption is reduced, making the lines more difficult to observe; but it is reduced less in the wings of a line than in the centre, giving the line an artificial breadth. The results of this paper support the assumption made in the theory that the line breadth due to collisions varies directly as the pressure, at any rate down to 0·03 mm. Hg. There is no reason to believe that this law will not be obeyed down to pressures of about 0·001 mm. Hg where the Doppler breadth would be of the same order as the collision breadth. To avoid disturbance of thermal equilibrium at such a pressure, the power entering a resonator, such as used in our experiments, would have to be reduced to about 10^{-10} watt. The power reaching a loosely coupled crystal detector would be about 10^{-12} watt, which is well above noise level if a superheterodyne method of detection is used. It should thus be possible to observe lines whose widths are about 10^{-6} cm.$^{-1}$

The phenomenon of disturbance of thermal equilibrium is not confined to the resonant-cavity method of measuring absorption, though it occurs at lower powers and higher pressures than in a waveguide transmission method owing to the higher energy density in the resonant cavity. The decrease in intensity of the ammonia lines observed by Good (1946) at low pressures in a waveguide has been attributed by the authors (Bleaney and Penrose 1946b) to disturbance of thermal equilibrium.

The derivation of the formula quoted there is given in Appendix B; the assumption was made, however, that the thermal relaxation time is equal to the mean time between collisions. The fact that it is found experimentally to be somewhat longer shows that the effect on the absorption will be rather more pronounced than calculated by the authors in that paper. The phenomenon of the disturbance of thermal equilibrium in waveguide measurements has been confirmed independently by Townes (1946), who shows that the effect is of the right order of magnitude if the thermal relaxation time is approximately equal to the collision time.

ACKNOWLEDGMENTS

The authors are indebted to Professor M. H. L. Pryce for considerable discussion and detailed criticism of the theory of this paper.

The work described in this paper was carried out partly on behalf of the Director of Scientific Research, Admiralty, and the authors wish to record their thanks for permission to publish. They also acknowledge gratefully a Senior Studentship from ths Royal Commission for the Exhibition of 1851.

APPENDIX A

Evaluation of (α/α_0) *for a cylindrical* H_0 *resonator*

In these experiments, a cylindrical cavity resonator was used excited in the H_0 mode, and the azimuthal component of the electric field at the point whose cylindrical coordinates are (r, ϕ, z) is

$$E = E_0 J_1(kr) \sin(2\pi z/\lambda_g), \qquad \ldots\ldots(9)$$

where k is determined from the relation $(kb) = 3{\cdot}832$, b being the radius of the cylinder, and λ_g is the wavelength in the cavity. The other components of the electric field are zero.

Insertion of (9) into (8) yields a double integral, to be evaluated over the coordinates r and z. The latter integration is simpler, being of the form

$$(\overline{n/n_0}) = A\int_0^{\pi}\{\sin^2\theta/(1+a\sin^2\theta)\}d\theta \qquad \ldots\ldots(10a)$$

where

$$a = \frac{4\pi^2}{3h^2}|\mu_{mn}|^2 E_0^2 J_1^2(kr)\tau\tau_e.$$

This integral may be evaluated explicitly, giving

$$\overline{n/n_0} = \frac{2}{a}\left(1-\frac{1}{\sqrt{1+a}}\right), \qquad \ldots\ldots(10b)$$

where the constant A has been determined by letting a tend to zero.

The remainder of the double integral is of the form

$$\overline{\overline{(n/n_0)}} = A'\int_0^b \frac{2\pi r J_1^2(kr)}{J_1^2(kr)}\left(1-\frac{1}{\sqrt{1+a'J_1^2(kr)}}\right)dr,$$

where

$$a' = \frac{4\pi^2}{3h^2}|\mu_{mn}|^2 E_0^2 \tau\tau_e.$$

Here the $2\pi r$ arises from the element of volume, and the $J_1^2(kr)$ in the numerator from the variation of the electric field strength in the numerator of equation (8). The integral may be written

$$\overline{\overline{n/n_0}} = A''\int_0^{3\cdot832} x\left(\frac{1}{1-\sqrt{1+a'J_1^2(x)}}\right)dx.$$

The constant A'' is determined by the relation (obtained by letting $a'\to 0$)

$$1 = A''\int_0^{3\cdot832} x\frac{a'}{2}J_1^2(x)dx$$

$$= A''a'0\cdot60.$$

Hence

$$\frac{\alpha}{\alpha_0} = \overline{\overline{n/n_0}} = \frac{1\cdot67}{a'}\int_0^{3\cdot832} x\left(1-\frac{1}{\sqrt{1+a'J_1^2(x)}}\right)dx. \qquad \ldots\ldots(11)$$

There does not seem to be any method of evaluating this integral except by numerical computation for different values of the parameter a'. This has been carried out for values of a' up to 200, at which the apparent absorption coefficient is smaller than the true value by a factor of about 20. The results are shown in figure 1, where (α/α_0) is plotted as ordinate and a' as abscissa.

It is convenient to relate a' to the total energy W' stored in the cavity. For a cylindrical H_0 resonator, whose length is m half-wavelengths, we have

$$W' = \frac{1}{8\pi}\int_0^{\frac{m\lambda_g}{2}} E_0^2\sin^2\frac{2\pi z}{\lambda_g}dz\int_0^b 2\pi r J_1^2(kr)dr$$

$$= E_0^2 V\,.\,3\cdot2\,.\,10^{-3},$$

where $V = \pi b^2\dfrac{m\lambda_g}{2}$ = volume of cavity.

Hence

$$a' = \frac{4\pi^2}{3h^2}|\mu_{mn}|^2\tau\tau_e\left(\frac{3\cdot1\,.\,10^2 W'}{V}\right).$$

It is obvious that similar computations may be carried out for any shape of resonator and mode of excitation, and that the results may be expressed in the form of a curve of (α/α_0) (similar to, though not identical with, that shown in figure 1) against a non-dimensional parameter a' which may be written

$$a' = \frac{4\pi^2}{3\hbar^2}|\mu_{mn}|^2 \tau\tau_e \left(\frac{gW'}{V}\right),$$

where g is a dimensionless constant which can be calculated from the field configuration in the cavity.

APPENDIX B

Evaluation of (α/α_0) *for rectangular* H_{11} *waveguide*

In a rectangular waveguide excited in the H_{11} mode the electric field has only one component, parallel to the narrow side, which varies as

$$E = E_0 \sin \pi x/b$$

from side to side, where b is the wide dimension of the guide. This leads to an integral similar to (10 a) above, with a solution

$$\frac{\alpha}{\alpha_0} = \overline{n/n_0} = \frac{2}{a}\left(1 - \frac{1}{\sqrt{1+a}}\right),$$

where

$$a = \frac{4\pi^2}{3\hbar^2}|\mu_{mn}|^2 \tau\tau_e E_0^2.$$

The power flowing down the guide is

$$W = \frac{\lambda}{\lambda_g}\frac{A}{2}\frac{\mathbf{c}E_0^2}{8\pi},$$

where A = cross-sectional area of guide,
λ = wavelength in free space,
λ_g = wavelength in guide.

Hence, substituting for τ in terms of $\Delta\bar{\nu}$, a becomes

$$a = \frac{16\pi|\mu_{mn}|^2}{3(\hbar\mathbf{c})^2(\Delta\bar{\nu})^2}\left(\frac{W\lambda_g}{A\,\mathbf{c}\lambda}\right)\left(\frac{\tau_e}{\tau}\right).$$

The equation for (α/α_0) can only be applied to a short length of guide, where W is substantially constant. If a long length of guide is used, in which there is considerable attenuation due to the absorption of radiation by the gas, a further averaging of (n/n_0) over the length of the guide must be carried out. No explicit expression can be given for this case, because the power will not vary exponentially along the guide when thermal equilibrium is disturbed. The attenuation will, in fact, be less near the input, where the energy density is greater, than near the output.

REFERENCES

Bleaney and Penrose, 1946 a, *Nature, Lond.*, **157**, 339 ; 1946 b, *Phys. Rev.*, **70**, 775 ; 1947 a, *Proc. Phys. Soc.*, **59**, 418; 1947 b, *Proc. Roy. Soc.*, A, **189**, 358.

Block, Hansen and Packard, 1946, *Phys. Rev.*, **70**, 127.

Fröhlich, 1946, *Nature, Lond.*, **157**, 478.

Good, 1946, *Phys. Rev.*, **70**, 213.

Purcell, 1946, *Phys. Rev.*, **69**, 681.

Purcell, Torrey and Pound, 1946, *Phys. Rev.*, **69**, 37.

Townes, 1946, *Phys. Rev.*, **70**, 665.

Van Vleck and Weisskopf, 1945, *Rev. Mod. Phys.*, **17**, 227.

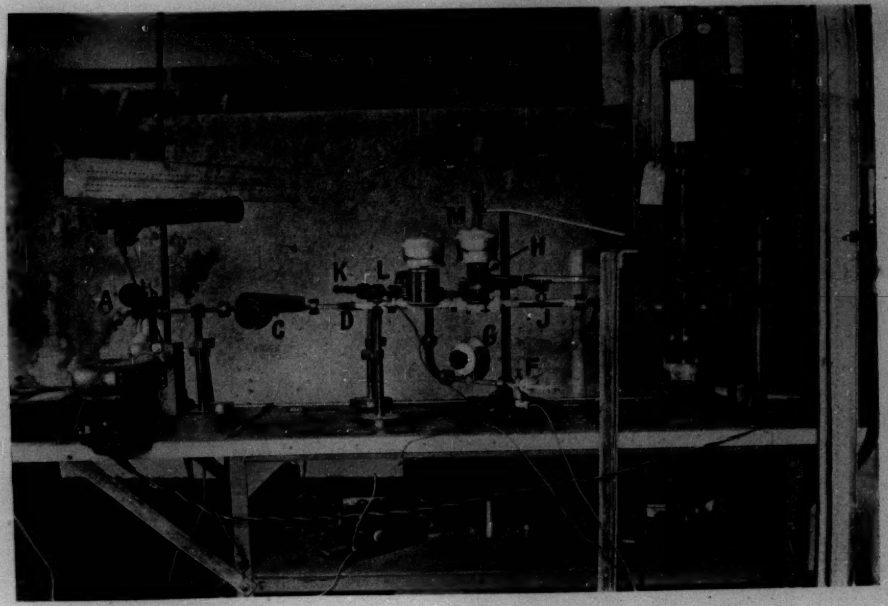


Figure 2. The apparatus.

A. Reflexion klystron oscillator.
B. Air blower.
C, G, J, L. Variable attenuators.
D. Waveguide "twist".
E. Cavity wavemeter.
F, K. Crystal rectifiers.
H. Resonant cavity (experimental chamber).
M. Micrometer head.

To face page 68

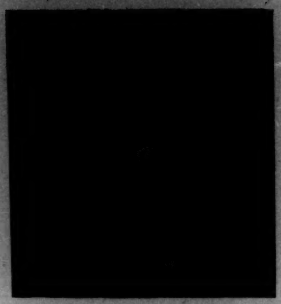


Figure 1 (*a*). Arc pattern with specimen perpendicular to beam.

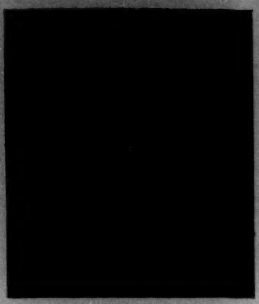


Figure 1 (*b*). Arc pattern with specimen inclined at 45° to beam.

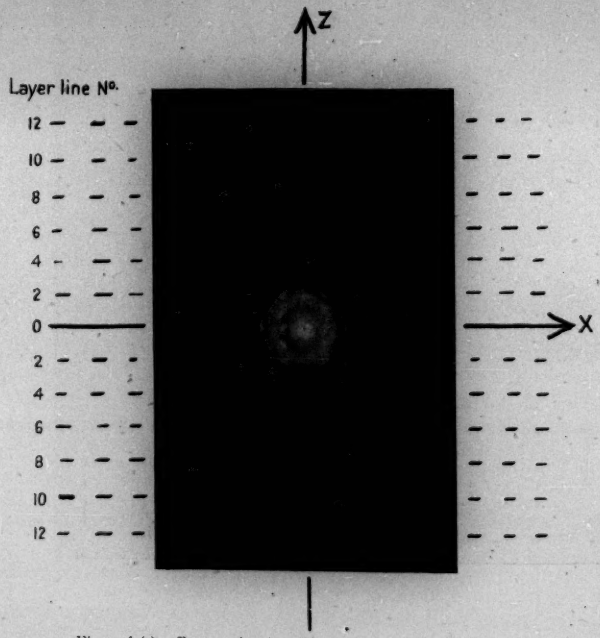


Figure 1 (*c*). Pattern showing diffuse zones in addition to arcs.

The Molecular Structure and Arrangement in Stretched Natural Rubber

By DOROTHY G. FISHER

Research Fellow of the Research Association of British Rubber Manufacturers at Imperial College

Communicated by G. I. Finch, F.R.S. ; *MS. received* 30 *May* 1947

ABSTRACT. Thin films of stretched rubber yield electron diffraction patterns which exhibit, in addition to the Laue spots, a pattern of diffuse zones.

Finch and Wilman (1937) obtained similar composite patterns from various organic crystals and, with Charlesby (1939), they showed that in the case of anthracene the diffuse zone pattern owed its origin to molecules in the crystal, but was similar to that which would be obtained by scattering from independent molecules oriented as in the crystal. The present study has demonstrated that this is also true in the case of rubber, and an appreciation of the significance of the diffuse-zone pattern has led to an advance in our knowledge of the structure of natural rubber.

§1. INTRODUCTION

X-RAY fibre diagrams were first obtained from stretched natural rubber by J. R. Katz in 1925. Since then many workers have examined rubber in both the stretched and unstretched state and have made estimates of the lattice constants of the pseudocrystalline structure produced on stretching. C. W. Bunn (1942) recorded an important advance when he arrived at a detailed atomic structure for the extended molecule by appreciating the relative intensity distribution in the x-ray pattern and applying the trial and error method. This structure, though it could not be claimed as final, did not conflict with the observed facts. Other workers, notably Jeffrey, with Cox and Bateman (1942, 1943, 1944), realising that the experimental data available from a fibre diagram are insufficient for a unique Fourier synthesis of the atomic positions, have preferred to examine the crystal structure of relatively simple materials containing isoprene units or related atomic groupings from which the configuration of the elementary units in the poly-isoprene chain might be inferred. It is clear, however, that we are still far from having reached finality in the determination of the molecular structure and arrangement in natural rubber.

Electron diffraction patterns have been obtained from both stretched and unstretched natural rubber (Trillat and Motz 1934, Kruilov 1935), but have hitherto added little to the data already provided by x-rays, chiefly owing to the fact that the intensity distribution in the electron diffraction patterns obtained from thin films is so markedly influenced by the relaxation of the third Laue condition.

In 1937 Finch and Wilman first drew attention to the fact that the transmission spot patterns obtained from single crystals of a number of aromatic compounds were associated with a remarkable pattern built up of diffuse areas of scattering,

and they recognised that this diffuse scattering pattern originated from the molecule and was not due to the lattice as a whole. Indeed, in 1939, with Charlesby, they showed that this diffuse-zone pattern was due to molecules which, though oriented with respect to the beam in accordance with the crystal orientation, were apparently executing thermal vibrations about their mean lattice positions. The pattern resembled, except for a slowly varying radial factor, that to be expected if the electron beam had been scattered by independent isolated molecules which had the same orientation as in the normal crystal. It is clear that this diffuse background pattern should, on account of its molecular origin, help in determining both the molecular structure and the molecular orientations in organic crystals which exhibit, as in anthracene, sufficient regularity of atomic arrangement within the molecule. It has now been found possible to obtain well-marked diffuse background patterns from stretched raw rubber, and these have thrown fresh light on the structure and arrangement of the molecules in natural rubber.

§2 EXPERIMENTAL

Transmission patterns were obtained from films which were stretched to about ten times their original length, the thickness of the film before stretching being of the order of 2000 A., as estimated from the interference colours under oblique illumination. The films were prepared by spreading a few drops of a solution of rubber in benzene on a clean water surface in a Langmuir trough; two chromium-plated bars were then placed across the trough and drawn apart, thus stretching the rubber film adhering to their lower surfaces. A portion of the stretched film was then picked up on a grease-free nickel gauze and examined in a Finch-type electron diffraction camera working with a camera length of 47 cm. and an accelerating voltage of 50–60 kv.

The types of pattern obtained are illustrated in figures 1 (*a*), (*b*), (*c*). Generally speaking all the patterns showed evidence of both the normal crystalline and the diffuse-zone patterns. The most favourable exposure conditions for bringing out the two were, however, different, a comparatively long exposure being required for the diffuse background pattern. Thus figures 1 (*a*) and (*b*) show the crystalline patterns obtained with a short exposure (about 10 sec.), with the film perpendicular to the beam for figure 1 (*a*) and inclined at an angle of 45° for figure 1 (*b*), the fibre axis remaining perpendicular to the beam in both cases, while figure 1 (*c*) shows a typical diffuse-zone pattern of long exposure (30 sec.) taken with the film perpendicular to the beam. Well-defined diffuse-zone patterns have not yet been obtained with the film sufficiently inclined to the beam to make any appreciable difference to this pattern, since the greater effective thickness of the film gave rise to sundry confusing effects such as a charging up of the specimen and loss of orientation under the prolonged electron bombardment necessitated by this greater thickness. It is also probable that the diffuse background pattern would still not be given prominently by the steeply inclined films, even if these disturbing influences could be eliminated, because in the case of a single crystal the intensity of the background pattern relative to that of the spot pattern should, according to Charlesby, Finch and Wilman (1939), vary inversely with the number of atoms in the crystal flooded by the electron beam, which is in this case proportional to the effective thickness of the film.

§3. THE ARC PATTERN

In appreciating the electron diffraction patterns the first step was to ascertain what measure of agreement existed between the arc pattern and the x-ray fibre diagram. A number of unit cells differing in detail have been proposed by various x-ray workers, but as C. W. Bunn (1942) has made the only complete structure analysis, his work was taken as giving the best available data; his proposed unit cell is monoclinic with $a = 12{\cdot}46$ A., $b = 8{\cdot}89$ A., c (orientation axis) $= 8{\cdot}10$ A., $\beta = 92°$, and space group $P2_1/a$.

It is immediately clear from the difference between figures 1 (*a*) and 1 (*b*) that the crystalline regions in the rubber films examined here have a strongly preferred orientation in the plane of the rubber film. A rough calculation indicated that this orientation was such that the *a* and *c* axes tended to lie in the plane of the rubber film. Working on this basis and using, for a first trial, the orthorhombic cell with $a = 12{\cdot}5$ A., $b = 8{\cdot}9$ A., $c = 8{\cdot}1$ A., $\beta = 90°$, which is a slightly simplified version of Bunn's proposed cell, it was found possible to index all the spots appearing in the present work (table 1). There were no results specifically requiring the monoclinic cell, but the possibility of this being the true cell cannot be ruled out, as a deviation of 2° in β would not in general be observable in view of the angular spread of the arcs caused by imperfections in the orientation. Bunn himself only introduced this deviation of β from 90° at a late stage in his calculations in order to account for one weak spot which is unlikely to occur in these patterns because the specimen was stationary during the recording of the pattern. The systematic absences, while conforming to the requirements for the space group $P2_1/a$, would equally well satisfy the higher and true orthorhombic symmetry of $P2_12_12_1$ requiring spots of types (h00), (0k0), or (00l) absent for odd values of h, k, or l; but again, reference to Bunn's calculated intensity values indicates that while (00l) spots are theoretically present for odd values of l they are so weak as to be unlikely to appear in the recorded pattern except in unusually favourable circumstances. The intensity distribution among the arcs in the electron diffraction patterns, figures 1 (*a*) and 1 (*b*), cannot be rigorously compared with that of the x-ray pattern, as it is influenced by the extent of the range of orientations present in the specimen, the relaxation of the third Laue condition, and the fact that the specimen was examined in selected stationary positions, thus enhancing the intensity from sets of planes which were preferentially oriented in a direction parallel to the beam. It is nevertheless of interest to note that, making some allowance for these factors, there is good qualitative agreement in intensity distribution between the electron diffraction patterns and the normal (00l) oriented x-ray fibre diagram (table 1). Thus it appears that, while the electron diffraction arc pattern would fit a rather simpler cell having the orthorhombic form and space group $P2_12_12_1$, there is no evidence which conflicts with the small change in β favoured by Bunn and the space group $P2_1/a$; furthermore, the intensity distribution is in qualitative accord with that of the x-ray patterns.

With regard to the absolute value of the identity period, however, a real discrepancy between the electron diffraction results and Bunn's values appears. Thus the best estimates of *a* and *c* measured against the graphite standard (110) ring, 1·228 kx. (Nelson and Riley 1945), were $12{\cdot}46 \pm 0{\cdot}02$ kx. and $8{\cdot}23 \pm 0{\cdot}02$ kx., compared with Bunn's values of 12·46 A. and 8·10 A. respectively, from which it is seen that the present estimate of the identity period differs significantly from

Table 1. Comparison of observed spacings with those calculated for orthorhombic cell having $a=12{\cdot}5$ A., $b=8{\cdot}9$ A., $c=8{\cdot}1$ A., and comparison of observed intensities with those found by Bunn.

Indices (hkl)	Electron diffraction patterns: Spacings			Electron diffraction patterns: Intensities observed		X-ray patterns: Intensities found by Bunn for (001) oriented fibre diagram	
	Calculated	Observed					
		Specimen perpendicular to beam	Specimen at 45° to beam	Specimen perpendicular to beam	Specimen at 45° to beam	Observed	Calculated
200	6·25	6·21	6·19	VS	—	S	64·5
120	4·19	4·18	4·14	MW	S	VS	254
400	3·12	3·08	—	M	—	M	15
*240	2·10	—	—	—	—	—	1
600	2·08						0
800	1·56	1·54	—	W	—	VVW	4·5
920	1·33	1·31	—	VW	—	VW	15
1000	1·25	1·23	—	W	—	VVW	7·5
111	5·40	—	5·35	—	VW	VW	2
201	4·95	4·92	4·89	VS	VW	S	24
121	3·72	—	3·69	VVW	VW	S	27
311	3·42	3·39	3·37	VVW	VW	MW	5
401	2·92	—	—	VVW	—	—	0
521	2·85						
*231	2·54	—	—	—	—	MW	3
611	1·97	—	—	VW	—		0·5
711	1·71	1·70	—	VVW	—	VW	1
801	1·53	1·51	—	VW	—	VW	1
002	4·05	4·05	4·05	VS	VS	MW	8
*012	3·69	—	—	—	—	MW	7
112	3·54	—	3·51	—	VW	MW	3
202	3·40	3·40	—	S	—	MS	26
*022	3·00	—	—	VVW	—	M	6
*122	2·91	—	—	—	—		0·5
312	2·76	2·75	2·73	VW	W	M	4
402	2·47	2·48	—	W	—	VVW	1
*322	2·43	—	—	W	—		2
412	2·38	2·38	2·36	VW	VW	MW	7
132	2·35						0·5
512	2·07	2·06	2·04	MS	VW	M	6
432	1·90	—	1·88	—	VW	M	2
242	1·86						0
*612	1·81	—	—	VW	—	VVW	2
113	2·53	—	2·50	—	VW	MW	11
203	2·48	2·51	—	VVW	—		1
*123	2·27	—	—	—	—	MW	6
323	2·02	2·04	—	VW	—	MW	3
403	2·04						
*133	1·97	—	—	—	—	MW	1
004	2·02	2·02	2·02	S	VS	MW	9
*124	1·82	—	—	—	—	MW	6
314	1·78	—	1·81	—	—		0
205	1·57	1·58	—	VW	—	No layer lines higher than 4th order mentioned	
506	1·19	1·17	—	VW	—		
516	1·18						
008	1·01	1·01	—	VW	—		

* Weak diffractions occurred on photographs from a specimen oriented with its plane at 60° to the beam.

† Values given here are rough means of hkl and $\bar{h}$kl values.

Bunn's value of 8·10 A. although the two values for the *a* axis are in good agreement. These measurements were made on strong arcs of the (h00) and (00l) type and the graphite comparison ring was introduced by permitting a drop of diluted graphite suspension in water to spread and dry out on the stretched rubber film. The third axial length *b* was not investigated in detail as no spots of the (0k0) type appear in transmission patterns on account of the strongly preferred orientation with this axis perpendicular to the plane of the rubber film.

Notwithstanding the difference in the absolute value of the identity period, the general agreement in form between the electron diffraction and x-ray results justifies the use of Bunn's atomic coordinates as a basis for a preliminary investigation of the molecular origin of the diffuse-zone pattern.

§4. THE DIFFUSE-ZONE PATTERN

The pattern reproduced in figure 1 (*c*) shows three distinct features: these are, (i) the haloes characteristic of the amorphous or liquid state, (ii) a pattern of arcs forming a typical crystal fibre diagram, and finally, (iii) a number of diffuse zones, the so-called diffuse background pattern ascribed by Charlesby, Finch and Wilman (1939) to molecular displacement by thermal motions. It is in these zones and their relation to the crystalline pattern that the main interest of this work lies.

Since the diffuse zones in the inner regions of the pattern are associated with clusters of arcs it might at first be thought that they arose from some distortion of the crystalline lattice giving rise to broadening of the arcs. Two considerations, however, indicate that this cannot be the whole explanation of the diffuse-zone pattern; these are that the broadening does not increase with increasing angle of diffraction, and that the diffuse zones occur in the outer regions of the pattern where there is no sign of associated arcs. Consequently we are led to some further consideration of the observed pattern in relation to the characteristics which would be expected on the molecular thermal vibration theory.

In connection with their work on anthracene, Charlesby, Finch and Wilman (1937, 1939) developed a theoretical expression for the intensity arising from a group of rigid molecules oriented in accordance with the crystalline arrangement, but undergoing thermal vibrations. They found that the resulting pattern should consist of two parts: (i) the Laue spot pattern characteristic of the periodicities of the crystal lattice, and (ii) a pattern of diffuse zones characteristic of the atomic arrangement within the molecules, and that the intensities of the diffuse zones should increase relatively to those of the Laue spots with increasing angle of diffraction. These conclusions were borne out by their diffraction patterns from anthracene, and the present rubber patterns show closely similar features, for it has been shown that the arc pattern corresponds to the normal crystalline arrangement, and it is readily apparent from figure 1 (*c*) that the diffuse zones become increasingly more intense relative to the arcs with increasing angle of diffraction, in fact, they remain prominent in the outer regions of the recorded pattern where the arcs have ceased to be even faintly visible.

Thus the evidence points strongly towards the conclusion that the observed diffuse-zone pattern has this molecular origin. In what follows it is assumed that this is the case and consequently that the arrangement of the diffuse zones is characteristic of the atomic arrangement within the molecules and should

therefore enable us to define the structure of rubber more closely than has hitherto been possible.

In order to use the diffuse-zone pattern in this way it is first necessary to consider the general form of the molecules involved and to make certain broad assumptions as to the manner in which they may be expected to vibrate. It is generally accepted that natural rubber is a long chain polymer of isoprene, C_5H_8, and that its rubber-like properties arise from the great flexibility which the molecules exhibit on account of the freedom of rotation of consecutive units about single bond linkages. It is therefore obvious that the theory developed for anthracene, in which the molecules were treated as rigid vibrating entities cannot be applied directly to the long flexible molecules of rubber; a fundamental treatment of the problem requires some account to be taken of the relative movements of the atoms within the molecule. As a beginning, however, the simplified case can be examined where the molecule is regarded as being divided into fairly short segments which undergo displacements relative to one another as a result of thermal vibrations while the atoms within a segment remain at rest relative to one another.

The question then arises as to what length of the molecule can be considered to be included in a single segment. One obvious choice would be to regard each isoprene unit as a virtually rigid vibrating entity. There would, however, be some considerable simplification in the work of interpreting the diffuse-zone pattern if it were supposed to arise from a longer length of the molecule, since such a segment, including a number of identical units, would show a greater degree of geometrical regularity. It therefore seems worth considering whether any such choice would be reasonable, and in this connection it is necessary to consider the arrangement of the molecules in an extended specimen of rubber. The stretched material is believed to contain a number of pseudocrystalline regions where sections of the chain molecules, oriented parallel to one another by the stretching, have linked together to form a three dimensional array; these regions are interspersed with disordered regions where the molecules are twisted and tangled in an entirely random manner. The length of a single molecular chain is about 10 000 A. (Staudinger and Bondy 1930), while the average linear extent of a crystalline region, as indicated by the breadth of the arcs in the X-ray fibre diagram, has been estimated to be of the order of 600 A. (Hengstenberg and Mark 1928). Thus a single chain is likely to pass through several regions of order and disorder. The effect on the fibre diagram of increasing the extension is to enhance its intensity relative to that of the halo pattern without in any way altering the crystalline spacings; thus it would appear that further extension causes more of the material to take up the ordered arrangement while leaving those regions which were already crystalline unaffected. On the other hand, the disordered regions are never entirely smoothed out however great the extension, as is apparent from the persistence of the halo pattern right up to the maximum extension which the material will maintain without fracture.

It would thus appear that we can legitimately regard the extended rubber specimen as comprising large numbers of comparatively rigid crystalline regions interlaced with disordered regions of greater mobility on which the continued extensibility of the material depends. From this argument it is apparent that there is some justification for supposing that the part of any chain molecule included in a single crystalline region might, to a first approximation, undergo

thermal vibrations as a rigid unit. Such a segment would have a definite repeat unit in the direction of the orientation axis (the identity period of the arc pattern), and could only give rise to diffractions along the same lines as the arcs in the crystal pattern. Reference to the pattern, figure 1 (*c*), shows that, except for a small angular spread similar to that observed in the arc pattern, the zones do in fact lie near these same lines, thus giving experimental support to the hypothesis. Therefore, in the discussion which follows, the diffuse-zone pattern has been treated as though it arose from molecular segments containing at least 10 to 15 repeat units and with a total length of some 100 A., each behaving as a self-contained array of scattering centres diffracting quite independently of any adjacent chain or any other segment in the same chain. It is nevertheless appreciated that once the general lay-out of the molecule has been defined some further refinements may be obtained by taking into account the relative movements of atom groups within a segment and considering the pattern to be anticipated from individual isoprene units. An investigation of this kind is now being undertaken.

The interpretation of the pattern has been approached by the trial and error method, and in order to do this it was first necessary to make some deductions, based on reasoning from other evidence, as to the atomic structure of the molecule. As mentioned above, rubber is a long chain polymer of isoprene, C_5H_8; the identity period along the extended chain is such that it must include two isoprene units, and Bunn (1942) has shown conclusively that it takes up the *cis* form in relation to the double bonds. By analogy with the long chain hydrocarbons it might be anticipated that the chain would be planar, but the identity period along the axis of elongation, which all observers agree to be in the region of 8·1–8·2 A., is too short to permit this without serious and improbable distortions of the accepted bond lengths and angles; consequently we are forced to the conclusion that the molecule is non-planar. There are, then, many ways in which the chain can be supposed to be kinked and twisted to give the observed identity period. For the present work, however, two models have been selected for investigation: the first is the structure proposed by Bunn which was taken because it was the only complete structure so far proposed on the basis of direct experimental results; the second, hypothetical, structure has been chosen as involving the minimum of distortion from the wholly planar form, and therefore forming a useful basic model on which further modifications can be imposed if necessary. It is a model in which all the carbon atoms of each isoprene unit are planar and the planes of consecutive units are parallel to each other and to the orientation axis; furthermore, all bond lengths and angles are normal, the reduction in identity period being attained by straining the connecting links between consecutive units out of the plane, so that the molecule forms a series of steps up and down as indicated diagrammatically in figure 2. This was, in fact, the molecule selected by Bunn as the initial basis for his calculations, but in order to satisfy his X-ray spot intensities he subsequently modified it considerably by introducing certain distortions in the bond lengths and angles and moving the methyl group out of the main plane of the isoprene unit. The atomic coordinates for one complete repeat unit of each structure are shown in tables 2 and 3, and their projections along three rectangular axes in figures 3 and 4.

The problem then is to find the positions of the intensity maxima arising from a chain segment comprising ten to fifteen such units. In computing such a pattern we have to take into account two factors: the form of the individual

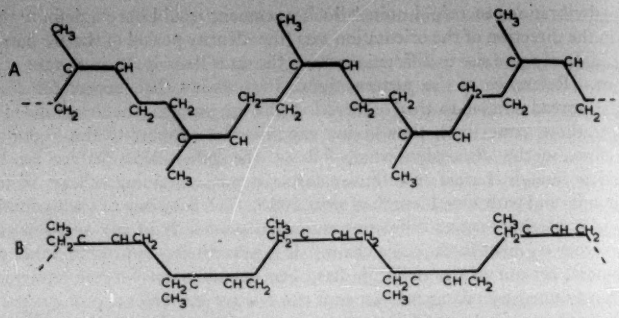


Figure 2. Rubber molecule—diagrammatic.
A. Projection in plane of double bands. B. Projection perpendicular to double bonds.

Table 2. Atomic coordinates (A), for Bunn's molecule—Single repeat unit

(*a*) As given by Bunn, *Proc. Roy. Soc.* A, 1942, **180**, 56. Left-handed "up" molecule. Referred to crystallographic axes $a=12{\cdot}46$, $b=8{\cdot}89$, $c=8{\cdot}10$ A., $\alpha=\gamma=90°$, $\beta=92°$.

	C_1 (CH_2)	C_2 (C)	C_3 (CH)	C_4 (CH_2)	C_5 (CH_3)	C_6 (CH_2)	C_7 (C)	C_8 (CH)	C_9 (CH_2)	C_{10} (CH_3)
x_1	0·753 *a*	0·854 *a*	0·845 *a*	0·745 *a*	0·968 *a*	0·744 *a*	0·644 *a*	0·659 *a*	0·757 *a*	0·532 *a*
y_1	0·899 *b*	0·865 *b*	0·905 *b*	0·959 *b*	0·876 *b*	0·834 *b*	0·874 *b*	0·905 *b*	0·834 *b*	0·828 *b*
z_1	0·802 *c*	0·703 *c*	0·542 *c*	0·457 *c*	0·773 *c*	0·326 *c*	0·215 *c*	0·052 *c*	0·025 *c*	0·268 *c*

(*b*) Referred to three orthogonal axes *x*, *y*, *z*, and to C_1 as origin. *y* and *z* are parallel to *b* and *c* respectively.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
x	0·000	1·259	1·147	−0·100	2·680	−0·112	−1·357	−1·172	0·050	−2·755
y	0·000	−0·303	0·052	0·533	−0·205	−0·577	−0·222	0·053	−0·577	−0·632
z	0·000	−0·846	−2·145	−2·792	−0·328	−3·851	−4·703	−6·034	−6·702	−4·224

Table 3. Atomic coordinates (A), for simple molecule having plane units and standard bond lengths and angles—Single repeat unit

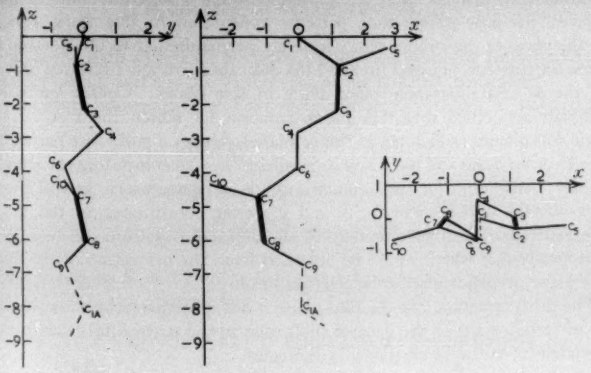
(*a*) Referred to orthogonal axes *x*, *y*, *z*, and to C_1 as origin.

	C_1 (CH_2)	C_2 (C)	C_3 (CH)	C_4 (CH_2)	C_5 (CH_3)	C_6 (CH_2)	C_7 (C)	C_8 (CH)	C_9 (CH_2)	C_{10} (CH_3)
x	0	1·26	1·26	0	2·51	0·04	−1·22	−1·22	0·04	−2·48
y	0	0	0	0	0	−1·22	−1·22	−1·22	−1·22	−1·22
z	0	−0·89	−2·22	−3·11	0	−4·05	−4·94	−6·27	−7·16	−4·05

(*b*) Referred to axes *x*, *y*, *z* as above, and to the midpoint of C_4-C_6 as origin.

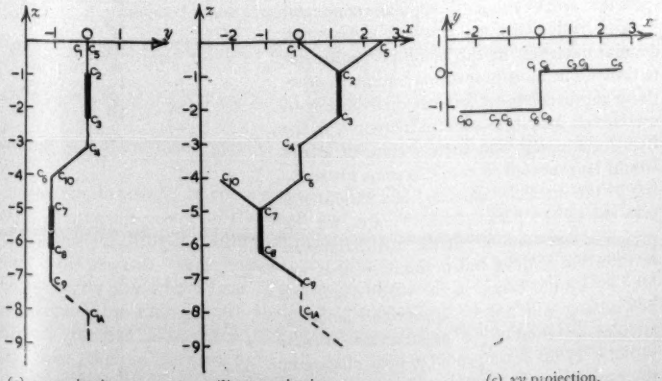
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
x	−0·02	1·24	1·24	−0·02	2·49	0·02	−1·24	−1·24	0·02	−2·50
y	0·61	0·61	0·61	0·61	0·61	−0·61	−0·61	−0·61	−0·61	−0·61
z	3·58	2·69	1·36	0·47	0·58	−0·47	−1·36	−2·69	−3·58	−0·47

unit, and the existence of a number of such units along the segment forming effectively a line diffraction grating with a definite identity period. It is only at those positions where both conditions are favourable that maxima can occur. The line grating alone would give rise to maxima along regularly spaced lines perpendicular to its own length, while the individual unit would give a more



(*a*) *yz* projection. (*b*) *xz* projection. (*c*) *xy* projection.

Figure 3. Rubber molecule proposed by Bunn. Projections along three orthogonal axes of one left-handed "up" molecule (Bunn's nomenclature); *z* axis in direction of extension.



(*a*) *yz* projection. (*b*) *xz* projection. (*c*) *xy* projection.

Figure 4. Suggested rubber molecule with planar isoprene units and standard bond lengths and angles. Projections along three orthogonal axes : *z* axis in direction of extension.

complex two-dimensional pattern. Consequently, in the computations which follow, the method adopted has been to select the lines along which the identity period permits maxima to occur and to examine the intensity variations which the atomic arrangement in the single unit gives rise to along these lines. In other words, the calculations are made as for a single unit, but the repetition

of this unit along the molecular segment introduces a geometrical regularity which limits the regions over which the resulting pattern need be considered.

Since the molecular segment is non-planar, it is no longer possible to construct a single two-dimensional intensity contour map adaptable to any crystal orientation by simple linear transformation of the axial lengths as was done in the case of anthracene; instead, each orientation requires a separate contour diagram. However, the only diffuse-zone patterns considered in this work were taken with the plane of the rubber film effectively perpendicular to the electron beam, and the arc patterns, figures 1 (*a*) and 1 (*b*), indicate a strongly preferred orientation with the *ac* crystallographic plane lying in this plane. Consequently we are principally concerned with this one orientation for which the *x* and *z* axes of atomic coordinates (which are in the *ac* plane), figures 3 and 4, are perpendicular to the electron beam and the *y* axis is parallel to it. In computing the anticipated intensity distribution for this orientation it is only necessary, as will be shown below, to take into account the *x* and *z* atomic coordinates of the molecule. More generally, for other orientations, the diffraction pattern can be considered as practically that which would be obtained from the projection of the molecule in the plane perpendicular to the electron beam.

The (010) orientation in the film plane is not quite perfect; thus the appearance of certain spots of the general (hkl) type in the pattern taken with the film perpendicular to the beam (table 1) indicates a partial rotation about the *c* axis. In addition, the *c* axis itself shows some angular deviation from its mean position within the *ac* plane as is indicated by the drawing out of all spots into arcs of about 10° angular range, and it seems likely that it will similarly suffer some angular deviation in the plane perpendicular to this. The most probable situation is something approaching a Gaussian distribution of orientation with its maximum frequency in the (010) orientation, and such a state of affairs would be expected to result in some broadening of the intensity maxima, but should not give rise to any peaks not anticipated for the preferred orientation. If the angular spread of direction of the *b* axis is similar to that exhibited by the *c* axis in the (010) plane the broadening will not be appreciable since an angular deviation of $\pm 5°$ about an axis normal to the electron beam produces little change in the projection of the molecule normal to the beam. In what follows, therefore, the pattern has been computed for the single orientation with the *xz* plane normal to the electron beam.

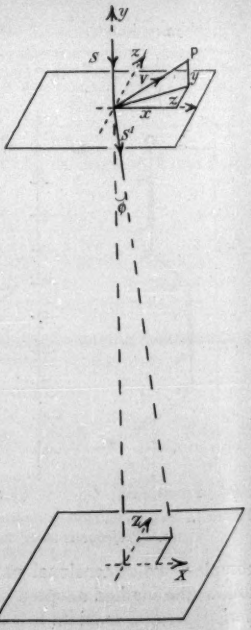


Figure 5. Illustration for derivation of intensity relationship.

The calculation of intensities is based on the following argument (see figure 5):—

Suppose the amplitude of the wave diffracted by any atom p in a direction ***S'***

from a wave incident in a direction $\boldsymbol{S}$ to be E_p (where $\boldsymbol{S}$ and $\boldsymbol{S}'$ are unit vectors). The intensity of the wave resulting from a group of such atoms is obtained by adding the amplitudes of the constituent diffracted waves, having due regard to their phases, and squaring the result.

Thus the resultant intensity is given by the expression

$$\{\sum_p E_p \cos(2\pi d_p/\lambda)\}^2 + \{\sum_p E_p \sin(2\pi d_p/\lambda)\}^2,$$

where d_p is the path difference between the wavelet scattered from p and from an arbitrary origin, and λ is the wavelength of the radiation considered.

For an atom whose position relative to the origin is given by the vector $\boldsymbol{v}_p$ we have

$$\begin{aligned} d_p &= \boldsymbol{v}_p \cdot (\boldsymbol{S}' - \boldsymbol{S}) \\ &= (\boldsymbol{x}_p + \boldsymbol{y}_p + \boldsymbol{z}_p) \cdot (\boldsymbol{S}' - \boldsymbol{S}) \\ &\simeq (\boldsymbol{x}_p + \boldsymbol{z}_p) \cdot (\boldsymbol{S}' - \boldsymbol{S}) \text{ since } \boldsymbol{y}_p \cdot (\boldsymbol{S}' - \boldsymbol{S}) \simeq 0 \end{aligned}$$

for small ϕ.

If the directions $\boldsymbol{S}'$ and $\boldsymbol{S}$ strike the photographic plate in points X, Z and 0, 0 referred to axes in the plate we have

$$\begin{aligned} \boldsymbol{S}' &= (\boldsymbol{X} + \boldsymbol{Z} + \boldsymbol{L})/\sqrt{X^2 + Z^2 + L^2} \text{ where } L \text{ is the camera length} \\ &\simeq (\boldsymbol{X} + \boldsymbol{Z} + \boldsymbol{L})/L \qquad \text{since } \phi \leqslant 2^\circ \end{aligned}$$

and $$\boldsymbol{S} = \boldsymbol{L}/L$$

so that $$(\boldsymbol{S}' - \boldsymbol{S}) = (\boldsymbol{X} + \boldsymbol{Z})/L.$$

Hence, the expression for the intensity becomes

$$\{\sum_p E_p \cos[(2\pi/\lambda)(\boldsymbol{x}_p + \boldsymbol{z}_p) \cdot (\boldsymbol{X} + \boldsymbol{Z})/L]\}^2 + \{\sum_p E_p \sin[(2\pi/\lambda)(\boldsymbol{x}_p + \boldsymbol{z}_p) \cdot (\boldsymbol{X} + \boldsymbol{Z})/L]\}^2$$

$$= \{\sum_p E_p \cos[(2\pi/\lambda L)(x_p X + z_p Z)]\}^2 + \{\sum_p E_p \sin[(2\pi/\lambda L)(x_p X + z_p Z)]\}^2 \quad \ldots\ldots(1)$$

when $\boldsymbol{X}$ and $\boldsymbol{Z}$ are taken parallel respectively to $\boldsymbol{x}$ and $\boldsymbol{z}$, and the axes are rectangular.

The amplitude E_p of the wave diffracted from an individual atom depends on the scattering power of the atom and on the angle of scattering, the latter relationship resulting in a falling off in intensity from the centre towards the outer parts of the pattern. With regard to the scattering power of the atoms concerned, the hydrogen atoms have been neglected as separate scattering entities, but have been regarded as adding a little to the effective weight of the carbon atom to which they are attached; thus the scattering factors of the C, CH, CH_2 and CH_3 groups have been taken in the ratio 6 : 7 : 8 : 9, which assumption was also made by Bunn in determining his atomic coordinates.

In computing intensities the following standard conditions have been taken: the factor λL has been taken as 2·75 A.cm. (for direct comparison with an experimental pattern the dimensions of the computed pattern have been reduced in the ratio L(observed)/2·75), and the origin of coordinates has been taken as coinciding with the carbon atom C_1 (figures 3 and 4). Atomic coordinates have been measured in Ångström units, and coordinates on the photographic plate in centimetres.

Thus, substituting for the various known or assumed values, the expression used in the computations of intensity at any point X, Z of the photographic plate becomes, for the Bunn molecule,

$$\left\{\begin{array}{l} 8\cos[(360/2{\cdot}75)(\quad 0 \quad - \quad 0 \quad)]\\ -6\cos[(360/2{\cdot}75)(\quad 1{\cdot}259X-0{\cdot}846Z)]\\ -7\cos[(360/2{\cdot}75)(\quad 1{\cdot}147X-2{\cdot}145Z)]\\ -8\cos[(360/2{\cdot}75)(-0{\cdot}100X-2{\cdot}792Z)]\\ -9\cos[(360/2{\cdot}75)(\quad 2{\cdot}680X-0{\cdot}328Z)]\\ -8\cos[(360/2{\cdot}75)(-0{\cdot}112X-3{\cdot}851Z)]\\ -6\cos[(360/2{\cdot}75)(-1{\cdot}357X-4{\cdot}703Z)]\\ -7\cos[(360/2{\cdot}75)(-1{\cdot}172X-6{\cdot}034Z)]\\ -8\cos[(360/2{\cdot}75)(\quad 0{\cdot}050X-6{\cdot}702Z)]\\ -9\cos[(360/2{\cdot}75)(-2{\cdot}755X-4{\cdot}224Z)] \end{array}\right\}^2$$

$$+\left\{\begin{array}{l} 8\sin[(360/2{\cdot}75)(\quad 0 \quad - \quad 0 \quad)]\\ -6\sin[(360/2{\cdot}75)(\quad 1{\cdot}259X-0{\cdot}846Z)]\\ -7\sin[(360/2{\cdot}75)(\quad 1{\cdot}147X-2{\cdot}145Z)]\\ -8\sin[(360/2{\cdot}75)(-0{\cdot}100X-2{\cdot}792Z)]\\ -9\sin[(360/2{\cdot}75)(\quad 2{\cdot}680X-0{\cdot}328Z)]\\ -8\sin[(360/2{\cdot}75)(-0{\cdot}112X-3{\cdot}851Z)]\\ -6\sin[(360/2{\cdot}75)(-1{\cdot}357X-4{\cdot}703Z)]\\ -7\sin[(360/2{\cdot}75)(-1{\cdot}172X-6{\cdot}034Z)]\\ -8\sin[(360/2{\cdot}75)(\quad 0{\cdot}050X-6{\cdot}702Z)]\\ -9\sin[(360/2{\cdot}75)(-2{\cdot}755X-4{\cdot}224Z)] \end{array}\right\}^2; \qquad \ldots\ldots(2)$$

and for the second, less distorted, molecule, a similar expression holds but with the appropriate atomic coordinates substituted for the Bunn coordinates.

With an identity period of 8·1 A. and with $\lambda L = 2{\cdot}75$ the "layer-line" separation is 2·75/8·1 cm.; in other words the lines near which our preliminary assumptions permit intensity maxima to occur have Z values (taken to the nearest $\frac{1}{2}$ mm.) of 0, 0·35, 0·7, 1·05, 1·35, 1·7, 2·05, 2·4, 2·75 etc. cm. Inspection of the photographs, however, suggests that the most pronounced maxima lie on the even-order layer lines (as might be expected from the similarity in form of the two halves of the molecule) and consequently the computations have been confined to those lines. They were carried out in general at 2 mm. intervals along these lines except in a few cases where intermediate values were obtained. Angles in the structure-factor expression were taken to the nearest degree.

The molecule has four possible orientations in the film, for the conditions governing its orientation are that the c axis direction (z axis of coordinates) shall lie in the direction of extension, and that the plane of the isoprene units (xz projection) shall be parallel to the plane of the film. From figures 3 (*a*) and 4 (*a*), which show the appropriate projection, these conditions are seen still to be fulfilled if the whole molecule is rotated through 180° about the x or z axis or both. Taking the axes of coordinates as fixed in relation to the film this is equivalent to changing the sign of either the x or z coordinates, or both. Now it is reasonable to suppose that any section of the film impinged on by the electron beam will include equal numbers of molecules in all the possible orientations; consequently in computing it is necessary to consider the intensities arising from each of these orientations, the resulting intensity at any point being the sum of that due to the separate constituents taking into account the basic assumption

of no phase relationship between adjacent molecules. In fact it is only necessary to consider two of the orientations because, owing to the squares, the expression (1) for the intensity arising from a single molecular unit is not affected when the signs of both the atomic coordinates are reversed. It is also only necessary to consider one quadrant of the pattern because the effect on the expression (1) of reversing the sign of either X or Z is precisely the same as that of reversing the sign of the corresponding atomic coordinates. In the case of the undistorted molecule the atomic configuration shows some degree of symmetry which is best seen from table 2 (*b*) in which the origin of coordinates has been moved to the midpoint of the repeat unit. This introduces a further simplification to the computations, for the two orientations x, z and $x, -z$ now give identical intensity distributions as long as we confine ourselves to the layer lines; for, provided that the ten atoms of the repeat unit are regarded as repeating indefinitely (which is what confining our attention to the layer lines implies), a change from the x, z to the $x, -z$ orientation merely implies a change of origin from the point C_1 to the point C_6 (table 3), which cannot have any effect on the computed pattern since the choice of origin is purely arbitrary. Therefore for the undistorted model calculations were carried out for the x, z orientation only, and the values were multiplied by two to compare with the summed intensities for the two orientations of the Bunn molecule.

§ 5. COMPARISON OF COMPUTED PATTERNS WITH OBSERVED PATTERN

The computed intensity distributions along the zero, 2nd, 6th, 8th, 10th and 12th layer lines are shown in figures 6 (*a*)–(*f*). The fourth layer line has been omitted as it showed no pronounced maxima. An attempt has been made to construct the patterns which would be anticipated from the positions of these maxima, and the results are shown in figures 7 (*a*) and 7 (*b*), together with a trace taken from an experimental pattern, figure 7 (*c*). In constructing these patterns it has been assumed that the visible breadth of the diffraction maxima corresponds to the half-maximum breadth of the intensity peak, and that the imperfections of molecular orientation are of the order indicated by the extent of the arcs in the Laue spot pattern, so that the whole pattern suffers an angular spread of about 9°. The length of the molecular segment considered has been taken as sufficient to limit the layer lines to true lines, the apparent breadth of the diffuse zones in the c axis direction arising solely from the small rotation referred to above.

The similarity in form between the computed and experimental patterns is readily apparent and gives ample support to the fundamental theory on which the calculations are based; that is, we are justified in assuming that the observed diffuse-zone pattern arises from molecules which are oriented much as in the crystal but behave as entirely independent groups of scattering centres. In other words, the pattern is characteristic of what would be expected from a group of oriented molecules in a pseudo-gaseous phase.

A close inspection of the curves (figures 6 (*a*)–(*f*)) reveals significant differences between the intensity distributions from the two models. The most striking difference is along the eighth layer line where the two curves of figure 6 (*d*) show entirely different characteristics, the Bunn model giving rise to three weak peaks at roughly 0·5, 1·5 and 2·5 cm. from the centre, while the undistorted model

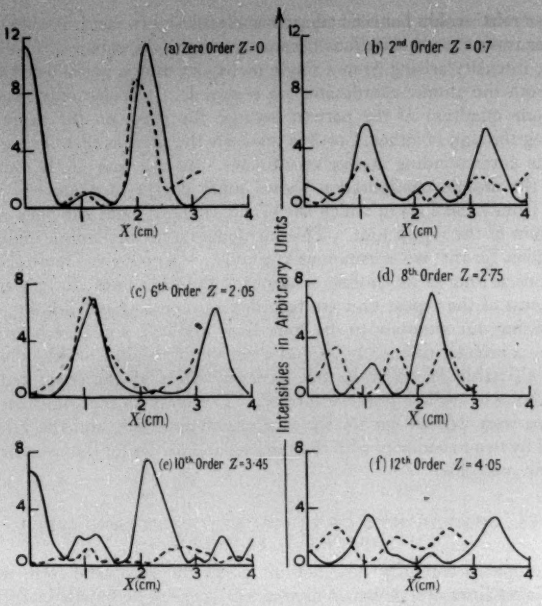
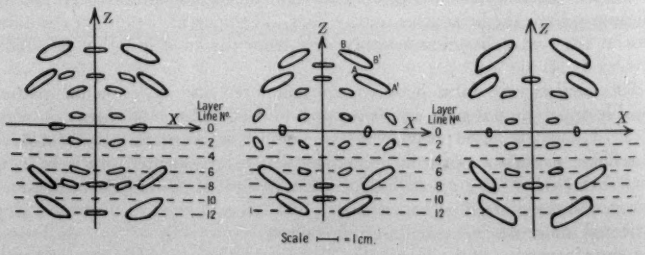

(a) Zero Order Z = 0
(b) 2nd Order Z = 0·7
(c) 6th Order Z = 2·05
(d) 8th Order Z = 2·75
(e) 10th Order Z = 3·45
(f) 12th Order Z = 4·05
Intensities in Arbitrary Units
X (cm)

Figure 6. Computed intensities along layer lines for the two atomic configurations investigated.
——— Model with planar units and standard bond lengths and angles.
– – – – Model due to C. W. Bunn.

gives a central strong peak, and a second of about the same intensity at 2·2 cm. Comparison with figure 1 (*c*), and with the traced experimental pattern, figure 7 (*c*), shows that the undistorted model is in far better agreement with experiment in this region; in fact, along this line the agreement is exact within the technical limitations of the patterns. The next characteristic difference



(*a*) Computed pattern for model due to Bunn.

(*b*) Computed pattern for model with planar units and standard bond lengths and angles.

(*c*) Observed pattern (traced from figure 1 (*c*)).

Figure 7. Computed and observed diffuse zone patterns.

between the two models is that the undistorted form gives sharper peaks; thus on the tenth and twelfth layer lines the Bunn model shows no pronounced peaks, but only certain wide zones of slightly increased intensity, while the undistorted model and the experimental pattern both show well defined peaks along these lines. The experimental data and undistorted model pattern are again in fairly good agreement in relation to peaks at 2·2 and 1·1 cm. on the tenth and twelfth orders respectively; these could well coalesce and give rise to the extended zone BB′ (figure 7 (*b*)), though the position of this observed zone is not quite what would be expected from the computed intensities. The error is probably not serious since the observed pattern is very faint and of poor definition in these outer regions, and the computed intensity maxima here are highly sensitive to small changes in the atomic coordinates owing to the large values of their coefficients. Along the zero and sixth order layer lines the two computed patterns are in good agreement with one another and with observation. On the second order line they differ again slightly, the undistorted model still giving more definite peaks though, while the peak at 1 cm. accords well with observation, the moderate peak to be expected at 3·2 cm. is absent from the experimental pattern.

It would thus appear that while both the molecular models investigated give rise to patterns agreeing in their broad general characteristics with those actually observed in the diffuse-zone electron-diffraction pattern, a more detailed examination shows that the computed intensity curves from the undistorted model are in much closer agreement with the facts than is the case with Bunn's model. This agreement is in fact remarkably good, the only discrepancy being the absence from the experimental pattern of a peak to be expected, in the case of the undistorted model, on the second order layer line at 3·2 cm. In particular, the undistorted model gives much the better agreement along the eighth order layer line, where the most sharply defined zones in the experimental pattern occur, and where Bunn's model gives an intensity distribution which is entirely at variance with observation.

§ 6. CONCLUSIONS

Although there are still some minor discrepancies to be resolved, the appearance of the diffuse-zone electron-diffraction pattern arising from thin films of stretched natural rubber is adequately explained on the basis of the thermal oscillations of the long-chain rubber molecules in the crystalline regions, this pattern being analogous with that which would be given by a stream of oriented molecules in a pseudo-gaseous condition. As Charlesby, Finch and Wilman (1939) pointed out the diffuse-zone pattern has distinct advantages in the examination of complex molecules, since it gives a direct indication of the configuration of the molecule independent of its mode of fitting into any particular crystalline lattice. In spite of the diffuseness of the patterns with which we are dealing a study of the diffuse zone or molecular pattern has shown that it is possible to discriminate decisively between two postulated atomic configurations which do not differ greatly.

In the case of natural rubber, neither of the two configurations investigated agrees in every respect with experiment, but the evidence brought forward above supports the view that the atomic arrangement in the stretched rubber molecule approximates to a simple form having planar units and standard bond lengths and angles, rather than to the more complex form hitherto considered.

ACKNOWLEDGMENTS

I wish to thank the Council of the Research Association of British Rubber Manufacturers for the grant of a R.A.B.R.M. Research Fellowship, which enabled this work to be carried out. I am also indebted to Professor G. I. Finch, F.R.S., for his valuable advice and encouragement, and to Dr. H. Wilman for many helpful discussions.

REFERENCES

BATEMAN, L. and JEFFREY, G. A., 1943, *Nature, Lond.*, **152**, 446.
BUNN, C. W., 1942, *Proc. Roy. Soc.*, A, **180**, 40.
CHARLESBY, A., FINCH, G. I. and WILMAN, H., 1939, *Proc. Phys. Soc.*, **51**, 479.
COX, E. G. and JEFFREY, G. A., 1942, *Trans. Faraday Soc.*, **38**, 241.
FINCH, G. I., 1941, *Proc. Roy. Soc.*, A, **179**, 67.
FINCH, G. I. and WILMAN, H., 1937, *Ergebn. exakt. Naturw.*, **16**, 353.
HENGSTENBERG, J. and MARK, H., 1928, *Z. Kristallogr.*, **69**, 271.
JEFFREY, G. A., 1944, *Trans. Faraday Soc.*, **40**, 517.
KATZ, J. R., 1925, *Chem. Ztg.*, **49**, 353; 1927, *Chem. Ztg.*, **51**, 381.
KRUILOV, K. I., 1935, *J. Exp. Theor. Phys.* (*U.S.S.R.*), **5**, 524.
NELSON, J. B., and RILEY, D. P., 1945, *Proc. Phys. Soc.*, **57**, 477.
STAUDINGER, H. and BONDY, H. F., 1930, *Ber. dtsch. chem. Ges.*, **63** (B), 724.
TRILLAT, J. J. and MOTZ, H., 1934, *C.R. Acad. Sci., Paris*, **198**, 2147.

Reviews of Books

Electronics and their Application in Industry and Research, edited by BERNARD LOVELL. Pp. xvi+660. (London: The Pilot Press Ltd., 1947.) 42*s.* net.

It is becoming trite to say that during the war the science of electronics developed at a phenomenal pace. Apparatus which, in 1939, was found only in the communications research laboratory, developed almost out of recognition to lead to the growth of an extensive industry concerned with the production of electromagnetic waves in the centimetre band, with the use of infra-red waves, and with the electronic control and measurement of industrial and medical processes. Here, indeed, is an example of the enormous rewards which accompany the devotion of the energy of great teams of scientists, and the coordination of work between Universities, Government laboratories, and technical industry to the theory, design and production of apparatus based on principles which previously were thoroughly understood and used only in the laboratory and by comparatively few workers. It is, however, remarkable that nearly all these advances have not been fundamental in nature, but technical in that they have been concerned with the practical application of electronic principles which had already been enunciated. It is those pre-war pioneers, mostly working in University laboratories, who gradually introduced order and logic to the understanding of the extraordinary versatility of the free electron in a gas at low pressure, to whom this remarkably interesting and informative book should be dedicated.

In a generous allowance of 660 pages there is described here, for the first time in a technical book, an advanced account of those aspects of electronics which developed apace during the war. Written by fourteen experts, the book provides a useful introduction to the theory and practice of the most recent developments in electronic apparatus. Each section is accompanied by an excellent bibliography, detailing the most important relevant papers so that the student can follow up the introduction to that aspect of modern electronics in which he is interested, and gain a reasonably complete knowledge of a particular field.

After an interesting introduction by the editor, Dr. Bernard Lovell, there follows an account of electron physics written by Dr. F. A. Vick. Though evincing that the author has a complete grasp of his subject, yet this section is inevitably condensed to such an extent that

the reader, unless he has already studied the subject in considerable detail, would fail to follow the arguments. It also seems unfortunate that this chapter is somewhat divorced from the rest of the book in that, once the basic electron emission equations have been established, they are not referred to again by any of the other authors: the standard of knowledge of modern physics demanded to appreciate Dr. Vick's contribution is much more advanced than that needed in any subsequent chapter. Indeed, a major criticism of this book is that there has been too little collaboration between the various authors so that there is no development of the subject as a whole as the book progresses. In view of the wide and varied nature of the complete text, this is to be expected to some extent, but one feels that closer links could have been established between the various sections, adding greatly to the book's value to the advanced student in giving a consistent picture of the development of modern electronic devices.

Photo-cells and television pick-up tubes are covered in two sections by authors from the research laboratories of Electric & Musical Industries Ltd., where most of the development work on iconoscopes and orthicons has taken place in this country. Drs. Lubszynski and McGee give competent accounts of their respective subjects, the section on television transmitting tubes being particularly valuable in that, for the first time in a technical book, the principles and practice involved have been properly described.

Dr. A. Elliott has contributed a fascinating chapter on photo-cells used in the infra-red which should prove of great value to those research workers who wish to use the highly sensitive devices developed during the war to establish secret signalling methods and to detect hot body radiation, and harness them for the study of the infra-red as used in spectroscopic analysis, astronomy, hygrometry and gas analysis.

The modern multi-segment magnetron, and the klystron tubes are described by Dr. F. C. Thompson in a chapter " Thermionic Valves for Very High Frequencies ". It seems a pity that the excellent account given of the action of these tubes could not be accompanied by extra matter dealing with their use in u.h.f. circuits, particularly since there is such scanty literature available on this subject. The same lack of circuitry is evident in a long section of the book devoted to Radar, written by Dr. R. A. Smith, though this omission is more than compensated by the interest which Dr. Smith stimulates in describing apparatus which, originally used for the investigation of the ionosphere, ultimately led to that extraordinary " box " which enables an observer in a ship or aircraft to obtain a fairly detailed picture on a cathode-ray tube screen of territory hidden by darkness or cloud.

The use of cold cathode valves for the control of industrial apparatus of certain types is described by Mr. L. Atkinson of Ferranti Ltd. Mr. H. Wood, from the same firm, contributes an up-to-date and valuable account of high frequency heating, and also describes an electronic apparatus for measuring moisture content.

Those highly ingenious devices which became known during the war as servo-mechanisms are the subject of a 50-page article by Mr. F. H. Belsey of Metropolitan-Vickers Ltd. Though difficult reading, this chapter is perhaps the best written section of the book; it deals with a complex subject in a logical and mathematical manner leading to a description of the types of gear employed.

In medical science, the applications of electronics are becoming increasingly numerous. Two chapters, " Electronics in Medicine ", and " Electronics in Physiology ", are contributed by Drs. Grimmett and Pumphrey respectively. The betatron, source of high-energy particles, is the subject of a special section by Dr. J. D. Craggs, who gives a theoretical and practical account of its development and use which clears up many misunderstandings and difficulties.

Finally, Dr. V. E. Cosslett, who is well-known for his research work and book on electron-optics, has written an excellent 100 page description of electron microscopy and electron diffraction.

This book, despite its size and the range of principles and techniques it describes, is singularly free from error; a tribute to the erudition of the various authors, to the editor, and to the excellence of the publisher's work. Unfortunately, as in nearly all technical books, the reader is occasionally obliged in one or two chapters to read the ill-constructed sentences of men who have devoted their lives to becoming technical experts. More than 400 excellent diagrams and photographs illustrate the text. It is a volume that every individual seriously interested in electronics should have on his bookshelf.

J. YARWOOD

New Developments in Ferromagnetic Materials, by J. L. SNOEK. Pp. viii + 136. (Amsterdam: Elsevier, 1947. Distributors for British Empire, Cleaver-Hume Press, London.) 13*s*. 6*d*.

The physical and mental frustration during five years of foreign occupation, with an "ever-growing burden of oppression and starvation" might well have dulled the urge to scientific enquiry in the Netherlands. That it did not do so is abundantly clear from this monograph, one of an extensive series on the progress of research in Holland during the war. The work of the author, a distinguished member of the Philips research group, on magnetic materials and much besides, was widely known and appreciated before the war; as was also his genial personality, which is reflected both in the general outlook conveyed by this book, and in the incidental and unobtrusive humour of word or phrase. The work here described is primarily that of the author and his associates at the Philips Laboratories between 1940 and 1945, but in the three main sections of the book the general background is sketched in sufficiently to indicate the setting of the various special investigations.

The first section, dealing with a number of themes under the general heading of the statics of ferromagnetism, includes an account of an experimental survey, by a comparatively rapid approximate method, of the crystal anisotropy and magnetostriction of binary and ternary alloys. A high initial permeability is associated with low values of these quantities, but even in those binary alloys for which zero values of each occur in a single phase range the zeros are unlikely to coincide at a particular composition. For a ternary alloy zero points become zero lines, and even a rough determination of the course of these lines shows whether unusually high permeability characteristics are likely for any composition. In the Fe-Si-Al system, for example, an intersection is indicated in the region where a maximum permeability had been found earlier by purely empirical methods. Although in the systems so far studied such intersections are infrequent, the investigation, which is still proceeding, has provided much valuable information.

Under the dynamics of ferromagnetism there is a discussion of eddy current effects, and of various after effects, tentatively classified as thermic, ionic and electronic. One long standing puzzle has been solved in that an entirely satisfying explanation, verified by experiment, is given of the often reported observations apparently showing that the initial magnetization curve crosses over the ascending branch of the hysteresis curve. The effect is a spurious one which may arise jointly from eddy current and demagnetization effects (it does not occur for ring specimens) to an extent depending on the rapidity with which the field is changed and the magnitude of the steps.

The more technical section of the book, on new materials, is mainly devoted to the extensive work which has been carried out on the cubic ferrites. A number of these ferrites, of the general formula $MO \cdot Fe_2O_3$, form mixed crystals, and four types of mixed ferrites have been developed, under the trade name "ferroxcube", as core materials suitable for high frequencies. The essential characteristic is a low conductivity combined, for certain mixtures, with a high initial permeability. For a mixed nickel-zinc ferrite, for example, a permeability of 4000 has been obtained. A particular point of interest here is that an enhanced permeability results from the addition to nickel ferrite of zinc ferrite, which is not itself ferromagnetic. Its effect is to lower the Curie point, so bringing the temperature region of high permeability into the working temperature range.

These few examples of the many topics touched on in this book must serve to illustrate its character and content. It is an informal account partly of completed work, and partly of work still in progress and of hypotheses which are being followed up. It should be in the hands of all those interested in ferromagnetism, by whom it will be valued both as a record of achievement under great difficulties, and for its stimulating presentation of new ideas.

E. C. S.